LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN
numero 58
giugno 1973
anno VI
volume X

Verso una stasi demografica mondiale

Oggi la popolazione del globo conta circa 3,6 miliardi di esseri umani. Un'estrapolazione dell'attuale andamento demografico mondiale indica la possibilità di stabilizzazione attorno agli 8,4 miliardi entro il 2100

di Tomas Frejka

Se una nazione desiderasse conseguire un incremento demografico nullo, in quale momento potrebbe prevederne la possibilità? Se i responsabili politici di un'altra nazione fossero preoccupati di un eventuale calo della popolazione, di quali dati potrebbero disporre per stabilire l'inizio probabile del declino? Per i prossimi decenni, quali alternative di crescita della popolazione totale o parziale del mondo sono da ritenersi realistiche? Che tipo di andamento demografico sarebbe opportuno ricercare, soprattutto a proposito della fecondità, in vista di un certo tasso d'incremento desiderato, cioè per ottenere una popolazione di una determinata entità?

Interrogativi di questo genere hanno costituito la base di un lavoro di ricerca al quale io stesso ho partecipato presso il Consiglio demografico, dopo averlo iniziato nell'Ufficio di ricerche demografiche della Princeton University. I principali strumenti metodologici sono state le proiezioni demografiche destinate ad agevolare l'interpretazione delle conseguenze a breve e a lungo termine dell'attuale andamento demografico di base, e ritenute indispensabili ai fini della valutazione di obiettivi realistici di variazioni demografiche. Tra le altre cose l'indagine ha dimostrato la necessità di accettare e programmare uno sviluppo considerevole della popolazione mondiale, rivelando come l'attuale rapporto di 30 a 70 tra popoli appartenenti a nazioni ricche e popoli appartenenti a nazioni povere sia destinato a passare inesorabilmente al rapporto di 20 a 80, se non addirittura di 10 a 90.

Fino al 1700 circa la popolazione mondiale era andata aumentando molto lentamente, probabilmente a un tasso medio inferiore allo 0,002 % l'anno. Oggi invece, sempre a livello mondiale, il tasso è mille volte superiore, essendo annualmente pari al 2 %. L'attuale stato di cose deriva da profonde modificazioni economiche e sociali, accompagnate per di piú da quel cambiamento radicale del quadro della fecondità e della mortalità che gli studiosi chiamano « transizione demografica ».

Per ogni popolazione si è soliti individuare tre stadi di trasizione demografica. Il primo è caratterizzato da tassi elevati e quasi pari di natalità e di mortalità, e quindi da un basso tasso d'incremento: la situazione che si è verificata per quasi tutti i popoli e quasi per l'intero corso della storia. Nel secondo stadio si ha una contrazione della mortalità, seguita a distanza di tempo da un'analoga diminuzione della fecondità: si determina cosí un elevato tasso di incremento demografico. Il terzo stadio è caratterizzato da bassissimi tassi di natalità e di mortalità e, quindi, da un tasso d'incremento gradatamente discendente.

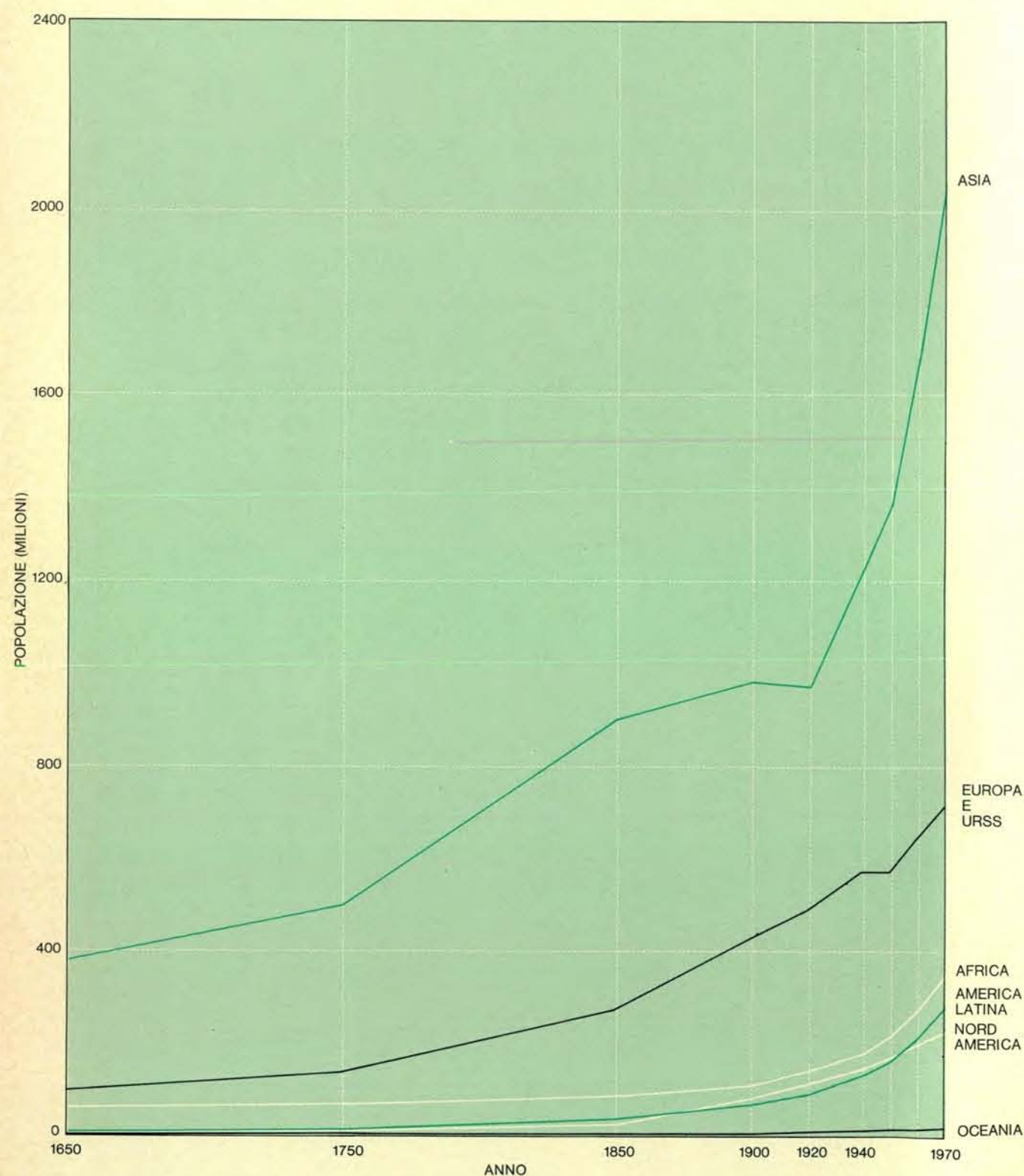
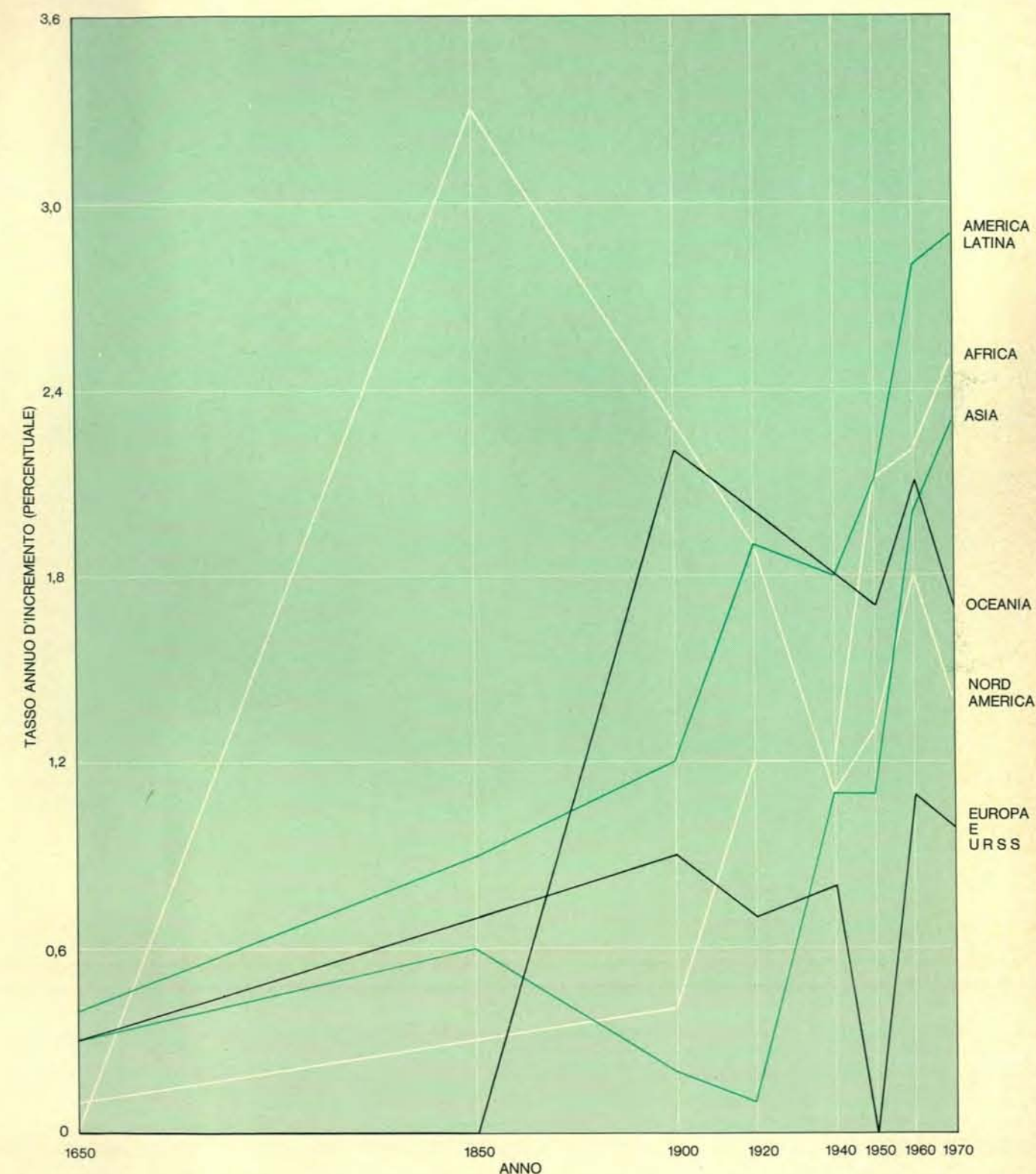
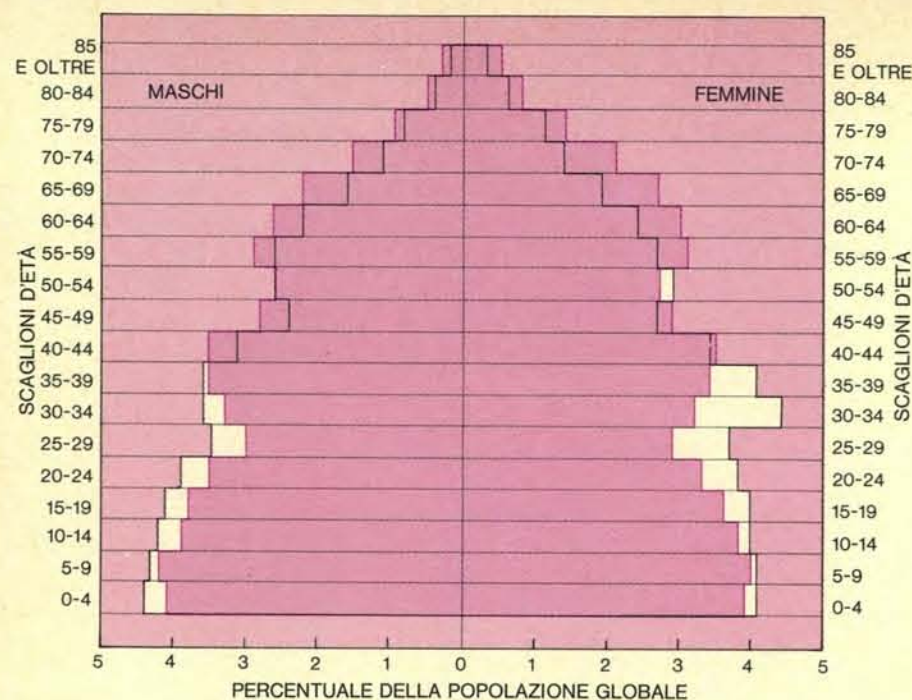
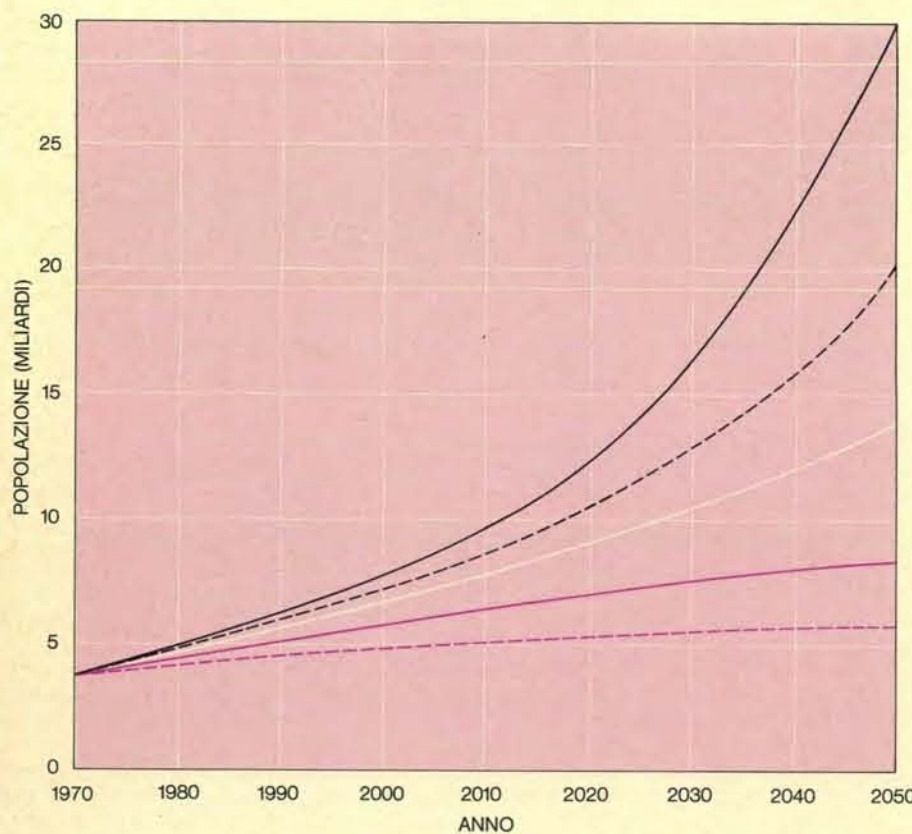
L'arco di tempo durante il quale avviene la transizione demografica varia notevolmente da popolo a popolo. Finora la transizione si è determinata essenzialmente nei paesi industrializzati, mentre i paesi in via di sviluppo stanno attraversando per il momento il secondo stadio. Inoltre in questi paesi, rispetto a quelli maggiormente sviluppati, la diminuzione della mortalità si è piú concentrata nel tempo, determinando tassi d'incremento di gran lunga superiori a quelli registrati nei paesi industrializzati. Questo è il motivo principale per cui complessivamente il tasso di crescita della popolazione mondiale è arrivato all'attuale livello annuale del 2 %.

L'esperienza storica fa pensare che prima o poi la popolazione del globo possa avvicinarsi allo stato di incremento nullo, cioè a una situazione in cui tutti i paesi si troverebbero nel terzo stadio della transizione demografica. Ma quali eventi si devono verificare affinché si giunga a questa fase? E quanto tempo ci vorrà, a prescindere da eventuali catastrofi di origine naturale o umana?

In primo luogo si deve tener conto del fatto che la persistenza dell'attuale tasso d'incremento (senza poi parlare di un aumento del tasso stesso) porterebbe entro l'anno 2000 la popolazione del globo a sette miliardi di individui. Il numero passerebbe a quattordici miliardi nel 2040 e a 28 miliardi nel 2070. (Attualmente siamo 3,6 miliardi).

In questi ultimi secoli però il comportamento demografico dei paesi più avanzati è andato mutando notevolmente: è cresciuto proporzionalmente il numero dei nuclei familiari di piccole dimensioni, mentre è andata aumentando notevolmente l'età media individuale. Una tale situazione della mortalità e della fecondità produce tassi d'incremento naturale (un dato che non tiene conto dei flussi migratori) inferiore annualmente all'1 %. La circostanza è dimostrata dal fatto che nei paesi più industrializzati il tasso d'incremento naturale continua a diminuire da una decina d'anni.

Ben diverso è il comportamento demografico dei paesi tuttora in via di sviluppo, ai quali le condizioni che presso le nazioni industrializzate avevano determinato livelli di mortalità decrescenti si sono comunicate con eccezionale rapidità, mentre non sono mutati affatto, oppure sono mutati lentissimamente, i livelli di fecondità. Sul finire degli anni sessanta del nostro secolo il tasso medio d'incremento naturale per i paesi in via di sviluppo è stato calcolato intorno al 2,6 % all'anno, con un tasso lordo di natalità di oltre 40 su 1000 abitanti e un tasso lordo di mortalità di circa il 15 ‰.

Nei paesi industrializzati, pur essendo stata accertata una certa variabilità da paese a paese nell'ultimo cinquantennio, è evidente che le popolazioni si riproducono in un regime di bassa mortalità e di bassa fecondità e che, anzi, tale regime presenta caratteristiche intrinseche prossime all'incremento zero. Eppure un esame più approfondito dei tassi lordi di mortalità e di natalità, nonché dei tassi d'incremento naturale, presentano un'apparente contraddizione: in effetti le popolazioni crescono a un tasso medio prossimo all'1 % all'anno. Il paradosso scaturisce dalla circostanza che le tendenze di crescita non sono determinate esclusivamente dai livelli di mortalità e di fecondità di una data popolazione; si deve infatti tener conto della struttura d'età, che è il risultato del precedente andamento della mortalità e della fecondità. I tassi lor-

Quadro dell'incremento regionale della popolazione mondiale dal 1650 al 1970. All'inizio del periodo considerato la popolazione complessiva ammontava a circa 553 milioni. Nel 1850 aveva toccato 1,3 miliardi. Dal 1900 è passata da 1,6 a 3,6 miliardi.

Le differenze regionali di incremento demografico balzano agli occhi quando si esprimano secondo il tasso di crescita gli aumenti di popolazione tra il 1650 e il 1970. Per migliaia di anni il tasso si era mantenuto sullo 0,002 %. Oggi è salito al 2 % circa.

La struttura d'età della popolazione nel Belgio (*piramide in colore*) e in Grecia (*piramide profilata in nero*) differisce e incide sul numero delle nascite e dei decessi. In Belgio il 13 % degli abitanti supera l'età di 65 anni; in Grecia tale valore è del 9 %. Pertanto il Belgio ha un più ampio settore della popolazione entro uno scaglione di età a elevato rischio di morte. Inoltre il Belgio ha un minor numero proporzionale di madri potenziali, giacché solo il 18 % delle donne si trovano nello scaglione d'età compreso tra i 20 e i 34 anni mentre per la Grecia questo scaglione rappresenta il 23 %.

Proiezioni diverse della popolazione mondiale in base a ipotesi diverse. L'incremento minore (*in basso*) si avrebbe qualora agli inizi del decennio in corso si fosse raggiunto un tasso netto di riproduzione pari a 1, il che implicherebbe una fecondità a livello di sostituzione. La popolazione continuerebbe ad aumentare per un certo periodo. Altre proiezioni partono da altre ipotesi: che tra il 2000 e il 2005 si raggiunga un tasso netto di riproduzione pari a 1 (*colore pieno*), che il medesimo risultato si abbia tra il 2040 e il 2045 (*bianco*), che proseguano gli attuali tassi di fecondità e di mortalità (*tratteggio in nero*) e che la fecondità rimanga costante con mortalità decrescente (*nero pieno*).

di, insomma, danno un'idea delle aggiunte e delle detrazioni relative alla popolazione attuale, ma non costituiscono una misura « reale » del livello di mortalità e di fecondità: sono infatti il risultato dei livelli di mortalità e di fecondità per ciascun scaglione d'età e del numero di individui corrispondente a ogni età.

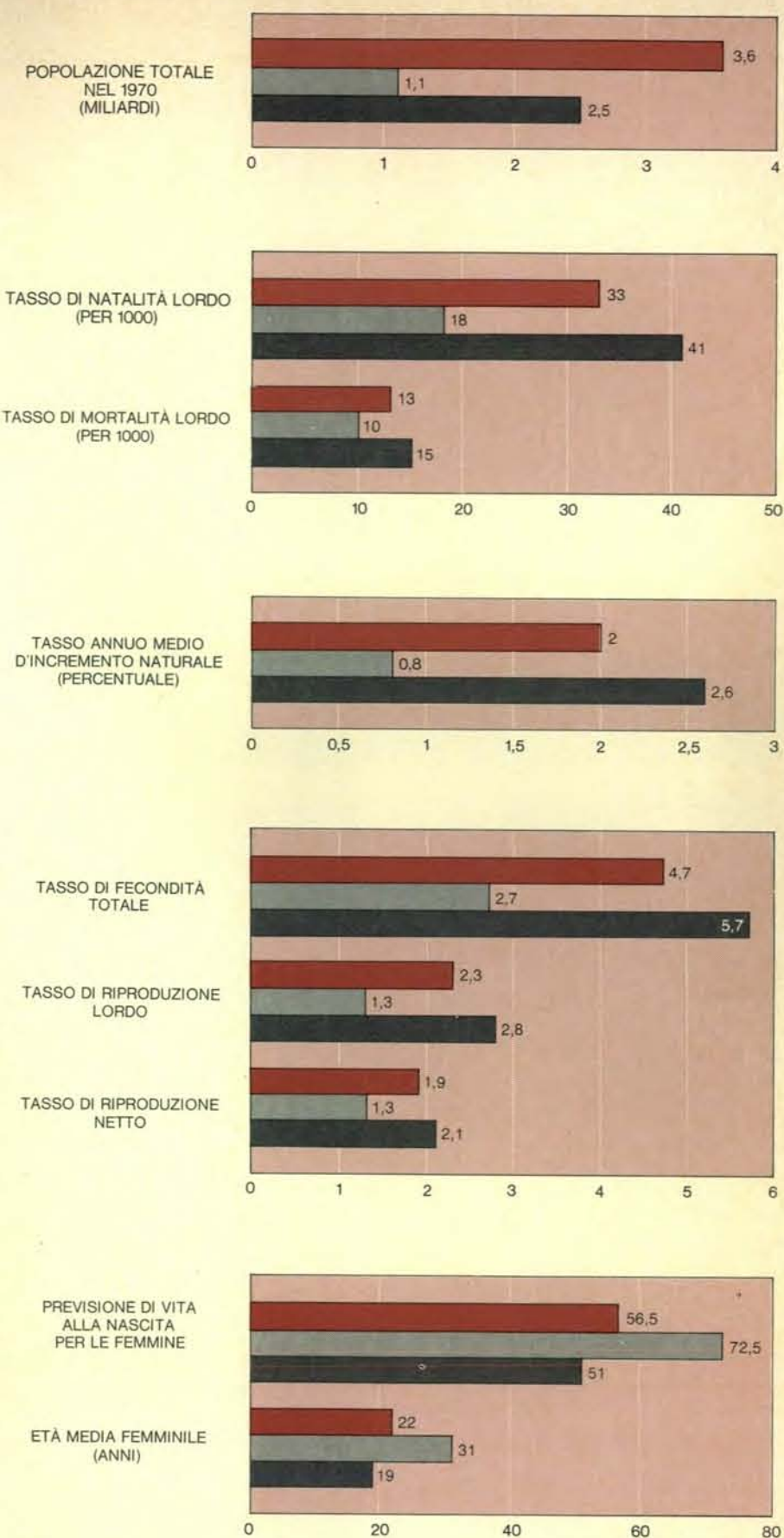
La Grecia e il Belgio costituiscono un chiaro esempio dell'incidenza della struttura d'età sulle tendenze d'incremento demografico. Nel 1967 il tasso lordo di mortalità è stato del 12 per mille in Belgio e dell'8,3 per mille in Grecia; il tasso di natalità lordo è stato del 15,3 in Belgio e del 18,7 in Grecia. Di conseguenza il tasso dell'incremento naturale è stato dello 0,3 % in Belgio e dell'1 % in Grecia. La differenza tra i tassi lordi dei due paesi era dovuta soprattutto alle differenze di struttura d'età (*si veda l'illustrazione a sinistra in alto*).

Tanto nell'uno quanto nell'altro paese si è riscontrato il medesimo andamento della mortalità (numero relativo di individui deceduti alle varie età) e della fecondità (numero medio di figli partoriti dalle donne dei vari scaglioni d'età). Differiscono invece notevolmente le strutture d'età, giacché in questo secolo la fecondità è stata più elevata in Grecia che in Belgio. Nel 1967 in Belgio si trovava una popolazione mediamente molto più anziana che non in Grecia (il 13 % della popolazione superava i 65 anni d'età, rispetto al 9 % per la Grecia), il che significa che in Belgio una parte notevolmente superiore della popolazione apparteneva ai gruppi d'età caratterizzati da un'elevata frequenza di mortalità. Al tempo stesso il Belgio presentava, rispetto alla Grecia, un minor numero di madri potenziali: gli individui di sesso femminile appartenenti allo scaglione d'età tra i 20 e i 34 anni ammontavano al 18 % della popolazione femminile, rispetto al 23 % in Grecia. Pertanto, mentre presso entrambe le popolazioni su 1000 madri potenziali nasceva il medesimo numero relativo di bambini, il numero assoluto dei neonati, al pari del tasso lordo di natalità, risultava più elevato in Grecia che in Belgio.

Certi criteri demografici alquanto più complessi rispetto ai tassi lordi di natalità e di mortalità forniscono maggiori informazioni sulla fecondità e sulla mortalità indipendentemente dalla struttura d'età della popolazione. Per la fecondità si ricorre al tasso di fecondità totale e al tasso lordo di riproduzione; per la mortalità si ricorre al metro rappresentato dalla previsione di vita alla nascita, mentre l'effetto combinato della fecondità e della mortalità si esprime mediante il tasso netto di riproduzione.

Il tasso totale di fecondità si esprime come numero medio di nati vivi che una donna partorirebbe negli anni di potenziale maternità, data la fecondità corrente delle donne a ogni età. In effetti la fecondità delle donne a una determinata età influisce sulle loro successive prestazioni riproduttive, tanto che nel 1973 la fecondità delle ventunenni (le donne che nel 1972 avevano vent'anni) può differire dal tasso accertato per le ventunenni nel 1972. Il tasso totale di fecondità è insomma un criterio artificiale riferito alla donna ipotetica ma basta a fornire con ragionevole approssimazione il numero medio di bambini partoriti da una donna all'interno di una data popolazione. Il tasso di riproduzione lordo si costruisce sulla medesima base teorica di quello di fecondità totale, ma si riferisce esclusivamente al numero medio di femmine che sarebbero partorite da una donna durante gli anni di potenziale puerperio.

La previsione di vita alla nascita è di circa settant'anni nei paesi industrializzati, il che implica bassi livelli di mortalità in ogni scaglione d'età; solitamente sono meno di 30 su 1000 i neonati che muoiono prima di compiere il primo anno di età, mentre circa il 95 % dei nati sopravvive almeno fino all'età di vent'anni. Tuttora in molti paesi in via di sviluppo la previsione di vita alla nascita non supera invece i quarant'anni, pur essendo molto ampia la gamma delle differenze per quanto riguarda la mortalità. Laddove la vita prevista è di circa quarant'anni, il quadro consueto presenta la morte di oltre duecento nati su mille prima del primo compleanno, mentre soltanto il 60 % circa di tutti i nati giunge almeno fino ai vent'anni.

Il tasso di riproduzione netto fornisce indicazioni sul numero di femmine che, nate dalla donna media, sopravviverebbero all'età feconda se fossero soggette all'andamento di mortalità corrente. Pertanto, in condizioni di mortalità elevata, la differenza tra tasso di riproduzione netto e tasso di riproduzione lordo risulterà notevole, mentre in condizioni di bassa mortalità il tasso netto è solo di poco inferiore a quello lordo. Tali relazioni sono opportunamente esemplificate dal confronto tra Costarica e Costa d'Avorio. Il tasso netto di riproduzione è più elevato in Costarica, mentre è più basso il livello di fecondità espresso dal tas-

Caratteristiche demografiche di tutto il mondo (*in colore*), dei paesi industrializzati (*in grigio*) e dei paesi in via di sviluppo (*in nero*). Il tasso totale di fecondità è una media dei bambini vivi nati da una donna. Il tasso di riproduzione lordo si riferisce al numero medio di femmine nate da una donna. Il tasso di riproduzione netto indica il modo in cui una generazione di donne feconde è sostituita da quella successiva. Un tasso di riproduzione netto pari all'unità comporta la sostituzione pura e semplice.

so di fecondità totale e dal tasso di riproduzione lordo.

Il tasso netto di riproduzione costituisce un indice di come si riproduca una data popolazione. Quando il valore è attorno all'unità, ciò dimostra che le condizioni correnti di fecondità e di mortalità sono tali da consentire all'attuale generazione di donne feconde di essere sostituita da una generazione di identiche dimensioni, sempre che permangano le condizioni accertate. La fecondità corrispondente a tale valore è indicata con il termine di « fecondità di sostituzione ».

Agli inizi degli anni settanta già diverse nazioni presentavano livelli di fecondità di questo tipo. Tra esse gli Stati Uniti, il Canada, la Danimarca, la Finlandia, la Svezia, la Bulgaria, la Cecoslovacchia, l'Ungheria, la Polonia, il Giappone e forse altri paesi. Ciò nondimeno le popolazioni di questi paesi continuavano ad aumentare, spesso a un tasso annuo prossimo all'unità percentuale, a causa della precedente storia demografica tuttora attiva per tramite della struttura d'età.

Proiezioni demografiche basate su cinque diverse ipotesi circa la probabilità di conseguire e conservare un tasso di riproduzione netto uguale a 1. Dal basso le cinque proiezioni corrispondono a: 1970-1975, 1980-1985, 2000-2005, 2020-2025 e 2040-2045. Se assumiamo un indice uguale 100 per la popolazione del 1970, gli indici corrispondenti salirebbero rispettivamente nel 2050 a 153, 172, 224, 287 e 357, e nel 2150 a 156, 176, 230, 306 e 416.

In quasi tutte le estrapolazioni demografiche di cui stiamo per occuparci si parte dall'ipotesi che sia stato raggiunto e conservato un regime di fecondità e di mortalità con tasso netto di riproduzione pari a 1. Tale valore non soltanto è giustificato per ragioni storiche (dal momento che il tasso di riproduzione netto deve essere rimasto intorno all'unità per migliaia di anni), ma anche perché qualsiasi altro valore che persistesse indefinitamente provocherebbe o l'incremento all'infinito o la riduzione a zero dell'andamento demografico. In tanti paesi industrializzati, inoltre, già si avvertono i sintomi di un comportamento demografico tendente appunto a un tasso netto di riproduzione attorno a valori prossimi a 1.

Nelle proiezioni demografiche da me preparate (una serie di cinque per ciascuna popolazione) il punto di partenza è costituito dalle caratteristiche demografiche attuali della popolazione in esame (*si veda l'illustrazione a sinistra*). Le proiezioni differiscono tra loro secondo il momento presunto in cui viene raggiunto, per essere poi mantenuto, il tasso netto di riproduzione pari all'unità. Differiscono quindi anche per il tasso presunto di attenuazione della fecondità. Si contempla inoltre l'ipotesi che l'attuale livello di fecondità scenda a un valore corrispondente a un tasso netto di riproduzione di 1 immediatamente (*proiezione 1*) oppure lungo periodi di 1, 30, 50 e 70 anni (*rispettivamente proiezioni da 2 a 5*). Si presume che il livello attuale di mortalità di ogni popolazione continui a scendere fino a stabilizzarsi sui minimi registrati finora. Per motivi tecnici i calcoli sono stati eseguiti esclusivamente rispetto alla popolazione femminile, ma i risultati sono approssimazioni ragionevoli delle caratteristiche e delle tendenze complessive di ciascun popolo.

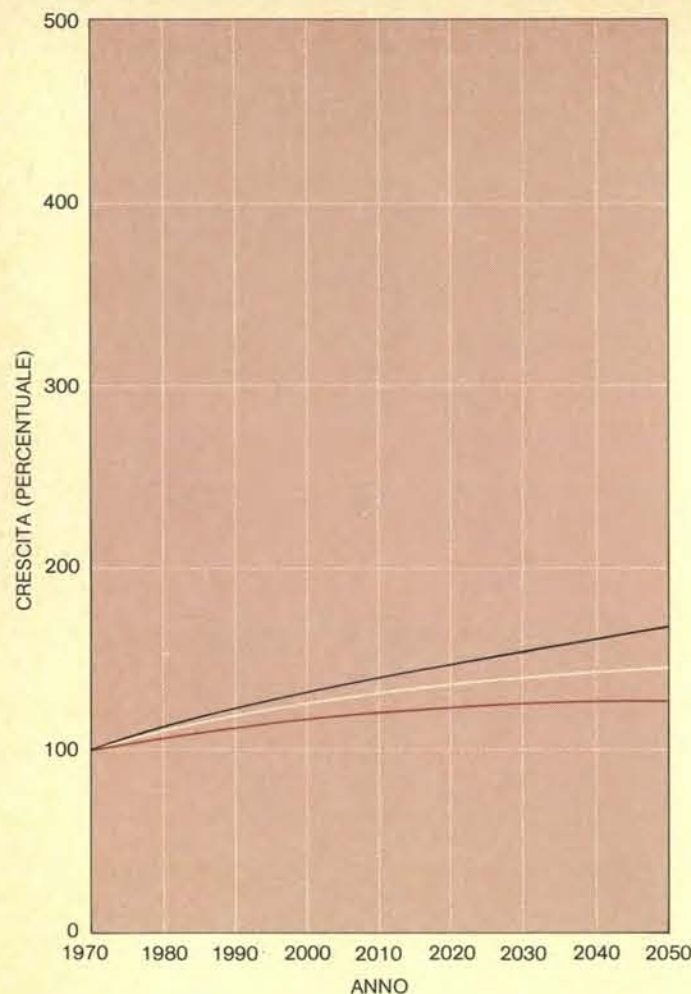
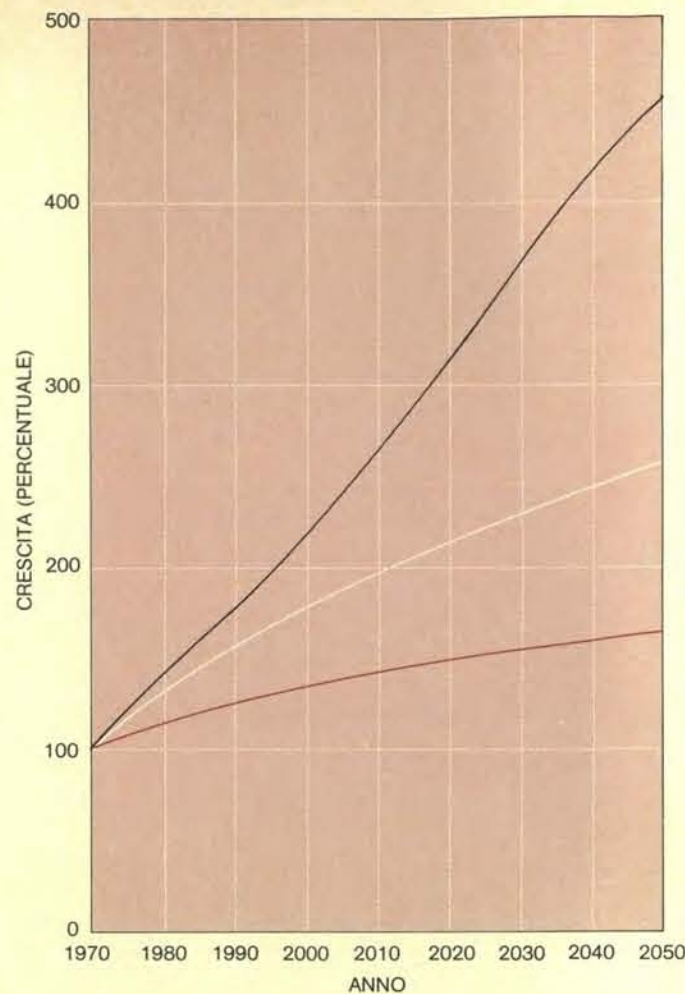
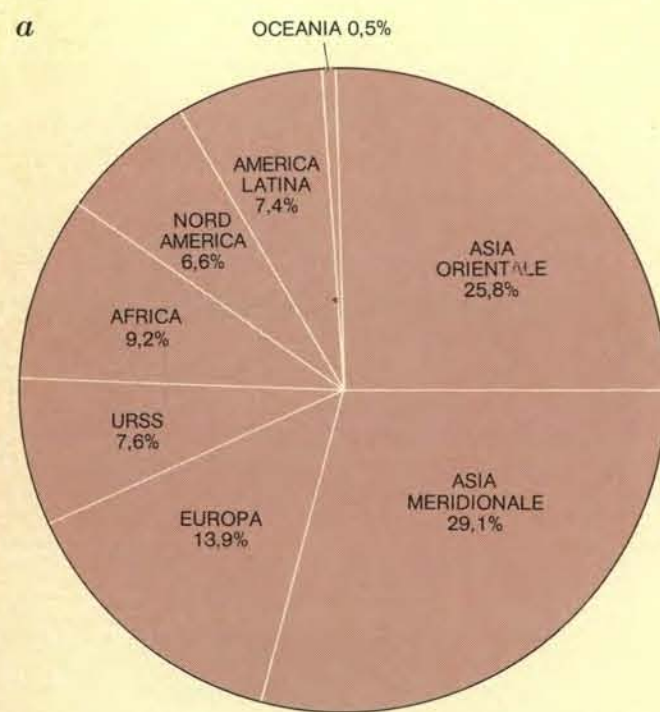
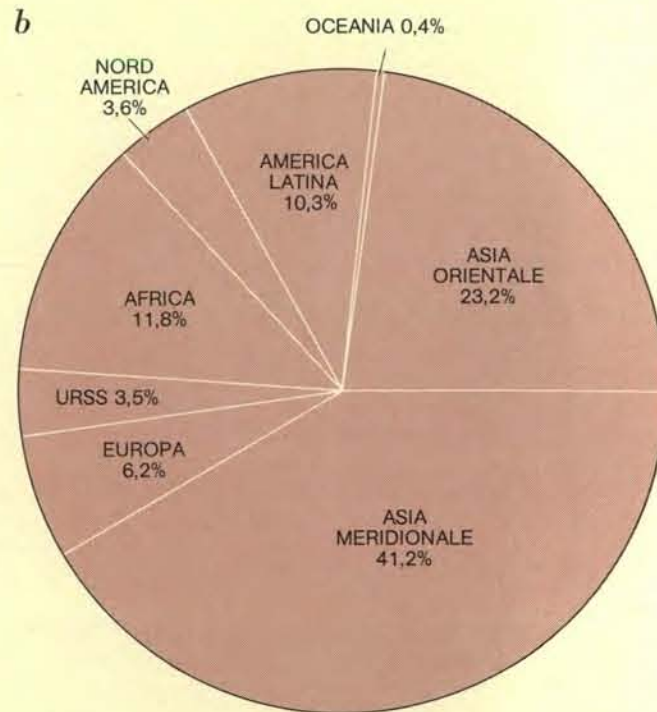
Il campo delle possibilità dell'incremento futuro della popolazione mondiale è talmente vasto da far ritenere opportuna la fissazione di limiti arbitrari quali valori estremi. Il minimo, che si raggiungerebbe solo se la fecondità diminuisse così rapidamente da far registrare già tra il 1970 e il 1975 un tasso netto di riproduzione pari all'unità, deve essere scartato come estremamente improbabile. Questo valore minimo alternativo offre comunque una preziosa indicazione, giacché esemplifica proprio quegli sviluppi che quasi certamente non si verificheranno mai. All'altro estremo si possono ipotizzare fecondità costante e mortalità decrescente. Tale alternativa, per quanto improbabile, non appare tuttavia impossibile, giacché rispecchia una situazione che di fatto esiste in buona parte dei paesi meno sviluppati.

All'estremo inferiore della scala lo eventuale immediato raggiungimento di un tasso netto di riproduzione mondiale pari all'unità determinerebbe sempre un incremento demografico del 30 % entro trent'anni e di oltre il 50 % entro ottant'anni (fino al livello di 4,7 miliardi nel 2000 e di 5,5 miliardi nell'anno 2050), giacché attualmente in quasi tutte le popolazioni del globo predominano gli individui in età feconda. A fecondità costante e a mortalità decrescente, dunque, la popolazione mondiale potrebbe salire a 7,4 miliardi nel 2000 e 30 miliardi entro il 2050.

	ANNO	POPOLAZIONE TOTALE (MILIARDI)	TASSO LORDO DI NATALITA	TASSO MEDIO D'INCREMENTO ANNUO	AUMENTO ANNUO MEDIO DELLA POPOLAZIONE (MILIONI)	PERIODO	TASSO DI RIPRODUZIONE NETTO	TASSO DI FECONDITA TOTALE
	1970	3,6	33	2,0	68	1965-1970	1,9	4,7
	SE NELL'ANNO	SI VOGLIONO OTTENERE I SEGUENTI VALORI				ALLORA NEL PERIODO	SONO INDISPENSABILI I SEGUENTI TASSI	
PROIEZIONE 1	2000 2050 2100	4,7 5,6 5,7	18 14 13	0,8 0,2 0,0	37 9 0	1970-1975 1980-1985 1990-1995 2000-2005	1,0 1,0 1,0 1,0	2,5 2,4 2,3 2,2
PROIEZIONE 2	2000 2050 2100	5,1 6,3 6,4	19 14 13	1,0 0,1 0,0	49 7 0	1970-1975 1980-1985 1990-1995 2000-2005	1,6 1,0 1,0 1,0	3,9 2,4 2,3 2,2
PROIEZIONE 3	2000 2050 2100	5,9 8,2 8,4	21 14 13	1,2 0,3 0,0	70 21 0	1970-1975 1980-1985 1990-1995 2000-2005	1,8 1,6 1,3 1,0	4,4 3,7 2,9 2,2
PROIEZIONE 4	2000 2050 2100	6,4 10,5 11,2	25 14 13	1,7 0,5 0,0	106 50 2	1970-1975 1980-1985 1990-1995 2000-2005	1,8 1,7 1,6 1,4	4,5 4,0 3,6 3,1
PROIEZIONE 5	2000 2050 2100	6,7 13,0 15,1	28 16 13	2,0 0,8 0,0	124 97 5	1970-1975 1980-1985 1990-1995 2000-2005	1,9 1,8 1,7 1,6	4,6 4,2 3,9 3,5

Questo è il quadro delle variazioni necessarie per conseguire taluni dei livelli di popolazione rappresentati nelle proiezioni a fronte. Il raggiungimento di livelli ancora maggiori richiederebbe un notevole decremento dell'attuale tasso di fecondità totale.

Poiché si tratta di limiti abbondantemente distanziati tra loro, le alternative sono numerose: l'andamento reale dipenderà dalle tendenze di mortalità e di fecondità. Nell'ipotesi che la mortalità continui a decrescere, è possibile prevedere che in corrispondenza di un declino sufficientemente rapido della fecondità la popolazione del globo possa restare contenuta all'incirca sui sei miliardi per l'anno 2000 (*proiezione 3*), mentre salirebbe a 6,7 miliardi in presenza di una fecondità solo moderatamente decrescente (*proiezione 5*). Nel primo caso la media attuale di 4,7 bambini nati da ogni donna in età puerperale dovrebbe scendere nel corso del prossimo trentennio a circa 2,2; nel secondo caso potrebbe essere di circa 3,5. Se si raggiungesse entro il 2000 e in seguito persistesse a livello mondiale una media di due nati in ciascuna famiglia (cioè a un tasso netto di riproduzione uguale a 1), la popolazione del globo, raggiunti gli otto miliardi circa nel 2050, non aumenterebbe piú in modo rilevante. Se si raggiungesse gradatamente un tasso netto di riproduzione pari a 1 entro la metà del prossimo secolo, allora la popolazione mondiale sarebbe di 13 miliardi nel 2050 e di circa 15 nel 2100.

Nonostante il fatto che le dimensioni future della popolazione mondiale siano predeterminate in modo notevole dalle attuali caratteristiche demografiche (livelli e andamenti di mortalità e fecondità, oltre alla struttura d'età), molto dipenderà dal variare di tali caratteristiche. A prescindere da eventuali catastrofi imponenti e persistenti, la mortalità è già abbastanza contenuta e probabilmente continuerà a decrescere. La fecondità è tuttora piuttosto elevata, essendo già a un livello doppio rispetto a quello che si determinerebbe in caso di incremento demografico zero, ma anche se diminuisse, è inevitabile che il decremento sarebbe graduale. Quanto alla struttura d'età, anch'essa è estremamente favorevole all'aumento, giacché solo una piccola parte della popolazione (meno del 6 %) supera og-

Prospetti di incremento demografico per i paesi sviluppati (*a sinistra*) e per quelli in via di sviluppo (*a destra*). Le prospettive sono quelle relative alle proiezioni 1 (*in colore*), 3 (*in bianco*) e 5 (*in nero*) riportate nelle illustrazioni precedenti.

Variazioni proporzionali della demografia rispetto alle otto principali regioni del globo, secondo le previsioni desumibili dalla proiezione 4, cioè dalla curva che ipotizza il raggiungimento di un tasso di riproduzione netto pari all'unità per il periodo tra il 2020 e il 2025. Le proporzioni attuali (*a*) sono messe a confronto con le proporzioni estrapolate per l'anno 2100 (*b*).

gi i 65 anni d'età, mentre una porzione rilevante delle donne si trova in età feconda. Ancora piú importante è il fatto che un numero relativamente elevato di donne entreranno in età feconda nei prossimi decenni, poiché attualmente circa il 37 % della popolazione globale ha meno di 15 anni. Inoltre, almeno durante il decennio in corso, a giudicare dagli attuali livelli e andamenti di fecondità e di mortalità, è molto probabile che nascano e sopravvivano generazioni di analoghe dimensioni di quella presente.

Se per la mortalità la tendenza futura dovesse restare piú o meno coerente con quella di questi ultimi decenni, in corrispondenza di un decremento abbastanza rapido della fecondità, può darsi che la popolazione mondiale si sviluppi secondo l'andamento rappresentato dalla proiezione 3: un po' meno di sei miliardi nel 2000 e un po' piú di otto miliardi nel 2050. A causa del potenziale di crescita intrinseco e giacché è irregolare prevedere in tutto il mondo una variazione del comportamento medio della fecondità tale da rendere possibile, su basi planetarie, la nascita di non piú di due o tre bambini per ogni donna entro il decennio 1980-1990, è da ritenersi estremamente probabile per l'anno 2000 un livello di appena cinque miliardi di individui.

L'illustrazione riprodotta a pagina 19 offre dei dati che possono dare una idea generale del tipo di sviluppo demografico che deriverebbe da diversi ritmi di decremento del tasso di fecondità. La proiezione 1, che dimostra come sarebbe necessario un tasso totale di fecondità di 2,5 già nei primi anni settanta affinché la popolazione mondiale potesse mantenersi al di sotto di 4,7 miliardi nel 2000 e di sei miliardi nel 2050, esemplifica, come si è accennato, degli sviluppi estremamente improbabili, se non impossibili. Per contro la proiezione 3, fondata sull'ipotesi di una caduta a 4,4 del tasso totale di fecondità nei primi anni settanta e a 2,2 entro il 2000, per avere una popolazione mondiale di 5,9 miliardi nel 2000 e di otto miliardi circa nel 2050, può essere considerata lo specchio di una serie di sviluppi che potrebbero verificarsi se la transizione demografica (soprattutto il decremento di fecondità nei paesi in via di sviluppo) fosse relativamente rapida e se si realizzasse entro una sola generazione il passaggio dalla fecondità di modello tradizionale a quella di andamento piú moderno. La proiezione 5, in base alla quale si è stabilito in 4,6 per i primi anni del decennio 1970-1980 e in 3,5 per il 2000 il tasso di fertilità totale, con una popolazione complessiva di 6,7 miliardi entro il 2000 e di circa 13 miliardi nel 2050, esemplifica invece gli sviluppi che potrebbero verificarsi se nei paesi in via di sviluppo la transizione demografica dovesse durare quanto è durata nell'Europa occidentale.

I dati citati forniscono una base per poter rispondere al quesito, che si pone molto di frequente, se non sarebbe possibile conseguire entro il 2000 un incremento demografico nullo. La risposta è che un evento del genere deve considerarsi estremamente inverosimile e probabilmente impossibile, a meno che non si verifichino variazioni inattese nei rapporti economici, politici e sociali mondiali, imprevisti sviluppi nei settori della scuola e della sanità o anche un numero senza precedenti di decessi. Per conseguire entro il 2000 uno sviluppo demografico nullo, inoltre, il tasso di fecondità totale in tutto il pianeta dovrebbe calare molto al di sotto del livello di sostituzione, per raggiungere all'inizio del nuovo millennio la quota di 1,1.

Nei paesi meno sviluppati, indipendentemente dalla rapidità del declino della fecondità, è probabile che nei prossimi decenni lo sviluppo demografico seguirà un andamento ben diverso da quello dei paesi industrializzati. Quand'anche non dovesse mutare l'attuale livello di fecondità, nel 2050 i paesi industrializzati non andrebbero al di là del doppio della loro popolazione di oggi, mentre nei paesi in via di sviluppo ci vorrebbe una riduzione drastica della fecondità per non superare nel 2050 il limite del doppio dell'attuale popolazione. E anche se entro il 2000 i paesi meno sviluppati raggiungessero un tasso netto di riproduzione pari all'unità, per il 2050 avrebbero già raggiunto un aumento di 2,5 volte. Con un decremento modesto e tuttavia significativo della fecondità, che verso il 2050 facesse scendere a 1 il tasso di riproduzione netto, la popolazione dei paesi meno industrializzati supererebbe sempre di 4,5 volte il valore attuale.

Nelle regioni a maggior grado di sviluppo è da prevedersi per l'anno 2000 un incremento demografico dal 15 al 25 %, mentre nel 2050 è probabile che in quelle stesse regioni si abbia un aumento dal 30 al 60 % rispetto alla popolazione del 1970. Tra oggi e l'anno 2000 è invece prevedibile che la popolazione dei paesi meno sviluppati si accresca dell'80 %, se non addirittura del 100 %. Qualora rimanesse costante l'attuale livello di fecondità e descrescesse la mortalità, entro il 2050 la popolazione dei paesi non industrializzati potrebbe risultare piú che decuplicata.

La classificazione socioeconomica adottata dall'ONU suddivide il mondo in otto aree principali: Asia orientale, Asia meridionale, Europa, Unione Sovietica, Africa, America settentrionale, America latina, Oceania. Le attuali condizioni demografiche differiscono da una regione all'altra. Sul finire del decennio scorso l'America latina aveva probabilmente il massimo tasso di crescita (2,9 % l'anno), una fecondità piuttosto elevata (tasso totale di fecondità pari a 5,5 dal 1965 al 1970) e un livello di mortalità relativamente basso (previsione di vita alla nascita per le femmine di 62,5 anni). L'Africa, pur presentando il piú alto tasso di fecondità totale (stimato in 6,4), registrò tuttavia un tasso annuo d'incremento demografico del 2,5 %, giacché in quel continente i livelli di mortalità sono i piú alti del mondo.

Evidentemente il maggior potenziale intrinseco di un ulteriore incremento demografico è quello offerto dalle vaste zone dell'Asia meridionale, dell'Africa e dell'America latina: la popolazione asiatica meridionale, per esempio, con un decremento della fecondità del tipo indicato dalla proiezione 5, potrebbe raggiungere verso la metà del prossimo secolo una consistenza cinque volte superiore a quella attuale.

La prevedibile e considerevole difformità nel ritmo d'incremento demografico da regione a regione si manifesterà con sempre maggior chiarezza nelle variazioni della distribuzione relativa della popolazione. Se, per esempio, l'accrescimento proseguisse secondo l'andamento della proiezione 4, la parte di popolazione che abita l'Asia meridionale passerebbe dall'attuale 30 % al 40 %, quella africana dal 9 al 12 % e quella latino-americana dal 7 al 10 %. Al medesimo titolo la parte di popolazione che vive in Europa scenderebbe dal 14 % al 6 %, quella dell'Unione Sovietica passerebbe dall'8 % al 4 %, mentre la popolazione nordamericana scenderebbe dal 7 a circa il 4,5 %.

Nei paesi demograficamente meno sviluppati, per i quali è unanime il parere sull'opportunità di un rallentamento del tasso d'incremento demografico, gli elementi-chiave saranno la data di inizio del declino della fecondità, la rapidità del declino stesso e, pertanto, il momento in cui saranno raggiunti livelli di fecondità che si avvicinino a quello di sostituzione. Per tali paesi varrebbe la pena di prendere congrue iniziative al piú presto in questa direzione.

Giordano Bruno

Generalmente si suppone che egli sia stato arso sul rogo per aver abbracciato il sistema copernicano. Pare però che le ragioni della sua adesione al copernicanesimo fossero più mistiche che scientifiche

di Lawrence S. Lerner ed Edward A. Gosselin

La mattina presto del 17 febbraio 1600, dopo otto anni di interrogatori e di reclusione, Giordano Bruno fu bruciato vivo sul rogo a Campo dei Fiori a Roma per ordine dell'Inquisizione: le accuse erano quelle di eresia e di apostasia monastica. Tra le altre colpe di Bruno erano l'adesione alla concezione copernicana del moto della Terra attorno al Sole e l'affermazione che la maggior parte dei corpi celesti, se non tutti, fossero popolati da esseri viventi, proprio come la Terra. Come è ben noto, l'oscurantismo ecclesiastico ridusse cosí in cenere uno tra i piú importanti precursori della scienza moderna. Nell'arco di una generazione le stesse forze avrebbero tentato di estinguere l'ultimo luminare del Rinascimento italiano: Galileo. Quando, tre secoli dopo, il Risorgimento restituí a Roma la luce e il sapere, gli intellettuali del mondo intero furono pronti ad acclamare in Bruno il primo martire della scienza nuova e un precursore dell'altro grande perseguitato, Galileo.

Non è difficile rendersi conto di come questo modo di considerare la figura di Bruno fosse una conseguenza del conflitto, costante nei secoli XVIII e XIX, tra la grande maggioranza degli intellettuali, di orientamento liberale e anticlericale, e la reazione pontificia. Si tratta nondimeno di una visione superficiale e del tutto erronea. Le sue facili generalizzazioni oscurano sia l'affascinante personalità di Bruno sia il suo importante contributo alle fondazioni filosofiche della rivoluzione scientifica. Da un'uguale cortina fumogena sono state avvolte le vere ragioni della sua condanna, cosí come la complessa connessione esistente tra la sua persecuzione e quella di Galileo.

Giordano Bruno, al secolo Filippo, nacque nel 1548 nella cittadina di Nola, nei dintorni di Napoli. All'età di 17 anni vestí l'abito di novizio domenicano e ricevette un'approfondita istruzione scolastica. Iniziò la sua carriera di studioso a Napoli, dove pare venisse richiesta la sua opera di insegnante nell'arte, allora di moda, della mnemonica. Fu a quest'epoca che cominciò a emergere la sua vena eterodossa. Nel 1576 fu costretto ad abbandonare Napoli come frate rinnegato sotto un cumulo di accuse di eresia; tra queste era quella di aver letto le opere proibite di Erasmo e di averle tenute nascoste presso di sé.

A quest'epoca le concezioni religiose e filosofiche di Bruno si erano già consolidate. Egli aveva abbracciato il sistema di fede del neoplatonismo ermetico, sul quale si fondò tutta la sua opera. Esamineremo piú avanti la concezione neoplatonica e la sua influenza sulla scienza fisica di Bruno. Per ora ci basti osservare che Bruno vide nel neoplatonismo una base per riconciliare cattolici e protestanti in un'epoca di violenta guerra religiosa.

Nel corso della sua fuga, Bruno pervenne a Roma, ma non vi rimase a lungo. Vagò poi per l'Italia, la Svizzera e la Francia, guadagnandosi da vivere mediante lezioni nell'arte della memoria. Pare che egli svestisse e rivestisse piú volte l'abito di frate. Conosciamo pochi particolari su questa parte della sua vita, ma sappiamo che, a causa della sua mancanza di tatto e della sua animosità polemica, lasciò dietro di sé, nella maggior parte dei paesi in cui soggiornò, sentimenti men che cordiali. Nel 1581 Bruno giunse a Parigi, dove si aggregò all'Accademia di Palazzo, un gruppo di letterati e filosofi alla corte di Enrico III.

Enrico viene spesso raffigurato come un re debole e indeciso, continuamente incerto tra la fazione cattolica ultraortodossa guidata dall'ambiziosa e potente famiglia dei Guisa e l'altrettanto potente fazione protestante guidata dalla famiglia di Borbone-Navarra. La vita personale di Enrico, e particolarmente il fatto che egli fosse un omosessuale e un travestito, non migliorarono la sua reputazione. E nondimeno egli lavorò – non del tutto invano – a evitare l'estrema disintegrazione di una Francia divisa e lacerata dalla guerra civile. Il suo scopo, nel fondare l'Accademia di Palazzo, fu quello di alimentare una base filosofica non dottrinaria per la riconciliazione politica delle fazioni cattolica e protestante. In quest'atmosfera il messaggio della riconciliazione religiosa di Bruno trovò un terreno fertile.

Nel 1583 Bruno ottenne lettere reali di raccomandazione per Michel de Castelnau, marchese di Mauvissière, che era ambasciatore di Enrico alla corte della regina Elisabetta. Bruno si recò a Londra e rimase al seguito dell'ambasciatore francese fino al 1585, quando il Castelnau fu richiamato a Parigi in conseguenza di mutamenti politici in patria. Durante il suo soggiorno inglese Bruno scrisse le sue opere piú importanti: sei diversi dialoghi. Uno di essi, *La cena de le ceneri*, contiene la sua esposizione piú diffusa della teoria copernicana.

Dopo il 1585 Bruno riprese i suoi vagabondaggi attraverso l'Europa, tenendo lezioni e scrivendo e riuscendo quasi sempre a mantenersi al centro di violente controversie. Egli continuò a sviluppare la sua « vera filosofia », che nelle sue speranze avrebbe dovuto costituire la base per una riconciliazione non solo tra cattolici e protestanti, ma anche tra uomo e Dio. Ritenendo che la trionfale assunzione al trono cattolico di Francia del tollerante Enrico IV, recentemente convertito, facesse presagire il compimento della propria missione, nel 1592 Bruno fece improvvisamente ritorno in Italia pensando di poter finalmente convertire il papa alla propria filosofia. Il suo anfitrione veneziano, Giovanni Mocenigo, lo denunciò all'Inquisizione locale, che dopo alcuni mesi lo consegnò all'Inquisizione romana. Otto anni dopo Bruno fu tirato fuori dalla sua cella e, dopo avere schernito i suoi carnefici e aver distolto gli occhi dal crocefisso che gli veniva presentato, fu dato alle fiamme.

L'Inquisizione non aveva avuto alcuna fretta nel liberarsi di Bruno, che non considerava una minaccia imminente. Di fatto la sua lunga prigionia fu in gran parte un colloquio macabro, in parte orale e in parte scritto, in cui ciascuna delle due parti cercava di convincere l'altra. È interessante osservare che l'Inquisizione dedicò ben poca attenzione alla difesa del copernicanesimo da parte di Bruno. Quando una volta Bruno sollevò il problema, sperando forse di portare fuori strada gli inquisitori, essi lasciarono cadere immediatamente l'argomento, passando invece alla questione delle relazioni di Bruno con Enrico III, Enrico IV e la regina Elisabetta. Pare inoltre che l'esecuzione di Bruno sia stata infine, in parte, una concessione del papa alla Spagna, la cui antipatia nei confronti di Bruno era molto piú forte e diretta che non quella del papato.

In varie occasioni Bruno affermò che la sua missione a Londra negli anni 1583-1585 era una missione ufficiale. Benché non ci siano prove incontrovertibili a sostegno di queste asserzioni, non c'è dubbio che gli obiettivi filosofici di Bruno fossero del tutto coerenti con i fini politici della monarchia francese. Nell'interminabile lotta a tre per la supremazia politica tra Spagna, Francia e Inghilterra, la Spagna aveva avuto la meglio. Era quindi naturale che la Francia cercasse di istituire piú stretti legami con l'Inghilterra al fine di ristabilire un certo equilibrio. La corona francese era però impegnata, almeno nominalmente, dalla parte dei cattolici nelle convulsioni che travagliavano l'Europa nel periodo seguito alla Riforma protestante. Un tale impegno politico era dunque una questione molto delicata. La cattolicissima Spagna si mantenne quindi pronta a spalleggiare alla prima occasione opportuna, una eventuale azione dei Guisa in Francia o della loro congiunta Maria Stuarda in Inghilterra.

Gli intellettuali cattolici liberali in Francia e la loro influente controparte protestante liberale tra i cortigiani inglesi cercarono con urgenza di determinare quali legami potessero essere forgiati tra loro. Un tale legame intellettuale avrebbe potuto diventare un efficace contrappeso all'obiezione dottrinaria a un'alleanza anglo-francese. Tra i cortigiani inglesi c'erano uomini del Rinascimento cosí poliedrici come il poeta filosofo sir Philip Sidney, lo astronomo-tecnico-mago e astrologo regio John Dee, il dotto copernicano Thomas Digges, il cortigiano-intellettuale-poeta sir Fulke Greville e William Gilbert, il medico regio, filosofo mistico e pioniere nello studio del magnetismo, che fu uno dei pochi uomini apprezzati da Galileo per i loro meriti scientifici. Fu in quest'ambiente che Bruno gravitò in Inghilterra ed è proprio in relazione a esso che dobbiamo valutarne gli obiettivi.

Come le due opere maggiori di Galileo, *La cena de le ceneri*, di Bruno ha la forma di un dialogo in lingua italiana. La maggior parte degli intellettuali alla corte di Elisabetta conoscevano bene questa lingua e la regina stessa era lieta di dimostrare la sua padronanza della stessa. Quel che Bruno si proponeva era niente meno che creare una comunità universale cri-

Il sistema tolemaico dell'universo (*metà superiore*), confrontato da Bruno con quello copernicano (*metà inferiore*) in *La cena de le ceneri*. Bruno si serví di questo diagramma per dimostrare l'ignoranza dei suoi avversari anticopernicani Nundinio e Torquato; in una vera discussione, avvenuta a Oxford nel 1583, Bruno fu sconfitto dai tolemaici soprattutto a causa della sua evidente ignoranza del sistema copernicano. Per esempio, nella metà inferiore dell'illustrazione, egli pone la Terra e la Luna, contenute da un medesimo epiciclo, in rotazione attorno a un punto immaginario orbitante attorno al Sole, invece di collocare la Terra in un'orbita attorno al Sole e la Luna in un'orbita attorno alla Terra. I simboli rappresentano, dall'esterno verso l'interno, le stelle fisse, Saturno, Giove, Marte, il Sole, Venere e Mercurio. Le xilografie originali erano in bianco su nero; sono riprodotte per gentile concessione della Harvard College Library.

stiana guidata dai suoi principi filosofici sotto la benigna protezione della regina Elisabetta. La posizione di lei, tollerante e ambigua in teologia, insieme ai suoi interessi intellettuali, ne facevano, agli occhi di Bruno, una candidata autorevole. In altra epoca pare che Bruno abbia guardato, come al protettore ideale di quest'iniziativa, a Enrico III e a Enrico IV.

Il dialogo ruota attorno al racconto di una tempestosa discussione che si suppone abbia avuto luogo a casa di Greville il mercoledí delle Ceneri del 1584. Il dialogo all'interno del dialogo vede in lizza il « Nolano » (cioè il Bruno stesso, nato, come si è detto, a Nola), impegnato in una difesa della teoria copernicana contro Torquato e Nundinio, due figure caricaturali di dotti protestanti ortodossi di Oxford, pedanti e sempre pronti a spaccare un capello in quattro.

Avviandosi alla casa di Greville, Bruno e i suoi amici si imbattono in tutta una serie di disavventure. Per cominciare, nessuno si presenta a riceverli all'ora fissata. Essi cadono in mano di due barcaioli fiacchi e grotteschi la cui barca geme acqua come un setaccio. Dopo aver varcato il Tamigi, si ritrovano in un vicolo dove sprofondano in una fanghiglia senza fondo e, dopo aver lottato a lungo col fango, finiscono col ritrovarsi non lontano dal punto di partenza. Si fanno poi strada attraverso una plebe bestiale e snaturata da infermità, che li spinge, li fa inciampare e li molesta. Quando arrivano alla casa di Greville vengono trattati con degnazione dai servi, finché riescono ad avere accesso al santuario dei cerimoniosi cortigiani elisabettiani. Con questo racconto Bruno intende simboleggiare la decadenza in cui la società è caduta in seguito al totale svisamento della vera religione e della vera filosofia.

L'illuminazione di una sfera (presumibilmente la Terra) da parte di un piccolo oggetto luminoso è uno degli esperimenti mentali di Bruno in *La cena de le ceneri*. Il piccolo corpo luminoso, se è collocato a breve distanza dalla sfera « opaca », ne illuminerà una piccola parte. Disposto a distanze successivamente maggiori (indicate con *b*) la sorgente luminosa illuminerà superfici sempre maggiori della sfera (*cd*, *ef*, *gh* e *ik*). Secondo Bruno, quando il « corpo lucido » sia allontanato all'infinito, sarà illuminata tutta la sfera. I punti sulla figura sono indicati con lettere diverse da quelle usate nel testo.

La discussione ha finalmente inizio. La natura della discussione e il carattere degli oppositori di Bruno risaltano con grande evidenza nella prosa ironica del Nolano. Teofilo, il portavoce di Bruno, descrive Torquato mentre si alza a parlare:

« Costui con un emfatico aspetto, col quale il *divum Pater* vien descritto nella *Metamorphose* seder in mezzo del concilio de gli Dei, per fulminar quella severissima sentenza contra il profano Licaone; dopo aver contemplato la sua aurea collana...; dopo essersi rizzato, ritirate le braccia da la mensa, scrollatosi un poco il dorso, sbruffato co la bocca alquanto, acconciatasi la beretta di velluto in testa, intorcigliatosi il mustaccio, posto in arnese il profumato volto, inarcate le ciglia, spalancate le narici, messosi in punto con un riguardo di rovescio..., appuntò le tre prime dita della destra insieme, et cominciò a trar di man dritti, in questo modo parlando: *Tune ille philosophorum protoplastes?* [Sei tu forse il primo dei filosofi?] »

Son questi i ridicoli oppositori anticopernicani di Bruno. È chiaro dalle osservazioni di Bruno che egli considera Copernico un grande matematico. Nondimeno, secondo Bruno, Copernico non ha compreso il vero significato metafisico che è alla base dei suoi calcoli fenomenologici. Comunicare il « vero » significato della teoria di Copernico è ovviamente il fine di *La cena de le ceneri*.

Come abbiamo ricordato, il dialogo mira a essere una difesa e un ampliamento della concezione copernicana del moto della Terra attorno al Sole contro l'opinione opposta, sulla quale era fondato il sistema tolemaico. Ma se si esaminano attentamente le argomentazioni del dialogo, ci si accorge con sorpresa che il copernicanesimo non è l'argomento vero dell'opera. Le argomentazioni scientifiche di Bruno, due delle quali esamineremo piú avanti, sono se interpretate letteralmente, insiemi di assurdità. Il fatto che il nostro giudizio possa godere del vantaggio a posteriori che ci è offerto dalle conoscenze scientifiche moderne non ha alcuna importanza. Quel che importa è che anche i contemporanei di Bruno avrebbero potuto rendersene conto se avessero stabilito di interpretare le argomentazioni alla lettera. Se i suoi nemici si fossero limitati a un'interpretazione letterale, avrebbero potuto coprire Bruno di ridicolo, ma difficilmente si sarebbero dati la pena di bruciarlo vivo.

Essendo evidente che le argomentazioni scientifiche di Bruno non erano destinate a essere interpretate alla lettera, qual era il loro senso vero? La soluzione di questo problema ci consentirà di comprendere la connessione esistente tra Bruno e la rivoluzione scientifica. Essa è vitale anche al fine di comprendere esattamente in che modo lo spirito di Bruno aleggiò come un incubo su Galileo nel vano tentativo da lui compiuto, tre decenni dopo, per convincere l'Inquisizione del fatto che le sue idee non erano né eretiche né politicamente pericolose.

Pare che Bruno abbia scelto deliberatamente la teoria copernicana come veicolo per introdurre le proprie concezioni filosofiche in parte perché essa godeva di una notevole corrente d'interesse e in parte perché poteva essere adattata alle proprie teorie. Bruno e i suoi contemporanei avevano inoltre ragione di vedere in Copernico un rinnovatore del pitagorismo magico, che essi consideravano una fonte della tradizione ermetico-neoplatonica. In primo luogo, i pitagorici avevano sostenuto una concezione eliocentrica; in secondo luogo, Copernico aveva alluso, nel *De revolutionibus orbium coelestium*, a Ermete Trismegisto, il presunto fondatore della tradizione ermetico-neoplatonica.

Le argomentazioni « scientifiche » di Bruno abbracciano un campo molto vasto. Egli affronta questioni astronomiche, fisiche, geologiche, archeologiche e geometriche, le quali sono però del tutto scientifiche in senso moderno e vanno valutate nei termini del posto che occupavano nel pensiero del tardo Rinascimento. Un aspetto del suo metodo è rivelato dai suoi esperimenti mentali di ottica. Bruno discute l'obiezione a Copernico che, supposta la Terra in movimento, col variare delle distanze fra la Terra e Venere la grandezza apparente di Venere dovrebbe mutare. Egli risponde che, benché la grandezza di un corpo non luminoso od opaco diminuisca con la distanza, ciò non è affatto vero per un corpo luminoso, poiché la visibilità di questo dipende dal suo splendore. Ne consegue quindi che dalla luminosità non può essere dedotta la distanza. Bruno schernisce i suoi oppositori, dimostrando l'assurdità delle loro argomentazioni mediante un argomento proprio non meno erroneo. Egli sostiene che se la distanza dall'osservatore a un oggetto luminoso (come una candela) venisse ripetutamente dimezzata, le dimensioni dell'oggetto aumenterebbero « sin tanto che vicinissimo venghi ad essere di quella grandezza che pensate ». Questa conclusione è chiaramente assurda. Bruno ignora il fatto che quest'argomento si applica tanto ai corpi luminosi quanto a quelli opachi.

Bruno dimostra che il Sole è piú grande della Terra. Se il Sole (*B*) fosse piú piccolo della Terra (*C*), allora il cono d'ombra (*DE*) proiettato dalla Terra sarebbe divergente. Se, al contrario, il Sole (*A*) sarà piú grande della Terra (*C*), il cono d'ombra (*HIF*) sarà convergente. Bruno ragiona cosí: « Sappiamo che l'ombra della Terra non eclissa mai Mercurio; perciò il Sole dev'essere piú grande della Terra ». Egli trascurava il fatto che sia nel sistema copernicano sia in quello tolemaico Mercurio si trova *tra* il Sole e la Terra. Se avesse usato Marte e non Mercurio, la sua argomentazione sarebbe stata corretta.

Stranamente, Bruno porta l'argomentazione erronea a una conclusione corretta: per un osservatore lontano, collocato su un'altra « stella », la Terra assumerebbe l'aspetto di un astro. La grandezza dei continenti oscuri diminuirebbe apparentemente con l'aumentare della distanza, mentre gli oceani, luminosi, conserverebbero il loro splendore. Infine, la Terra apparirebbe totalmente essere una stella. Inversamente, Bruno conclude che molte stelle sono essenzialmente simili alla Terra e solo la loro grande distanza ci impedisce di renderci conto che sono composte di terra e d'acqua.

Bruno realizza quest'argomentazione in modo molto curioso. Un piccolo corpo luminoso che si trovi a breve distanza da una sfera maggiore non luminosa illuminerà una piccola parte della sfera (*si veda la figura nella pagina a fronte*). Col crescere della distanza tra i due corpi, verrà illuminata una parte piú grande del corpo maggiore, tanto che infine sarà illuminato un intero emisfero. Ora, dice Bruno, aumentiamo ancor piú la distanza. Sarà allora illuminata una superficie ancora maggiore e se la distanza diventa infinita sarà illuminato l'intero corpo non luminoso!

Bruno procede poi a un'ulteriore generalizzazione. Consideriamo due piccoli corpi luminosi tra i quali sia un corpo maggiore opaco, cosicché la luce emessa da ciascuno dei due corpi luminosi non possa raggiungere l'altro. Se la distanza tra i corpi viene accresciuta indefinitamente, il corpo opaco si ridurrà a nulla, mentre i due corpi

Un corpo opaco posto tra due corpi luminosi piú piccoli non impedirà il passaggio della luce dall'uno all'altro se i due corpi luminosi saranno sufficientemente lontani. Man mano che una delle due sorgenti viene spostata a distanze sempre maggiori dal corpo opaco (*cerchio minore*), l'angolo alla sorgente diventa sempre piú acuto. A una distanza sufficiente *K*, cosí sostiene il filosofo nolano, « bisogna ch'al fine dovenghi... [l'angolo] l'acutissimo, et al fine a quel termine, che non appaia piú angolo, ma linea... Là onde al fine è necessario che dui corpi piú luminosi... non saranno impediti per non vedersi reciprocamente ».

luminosi conserveranno la stessa grandezza. A una distanza sufficiente il corpo opaco non impedirà piú il passaggio di luce tra i corpi luminosi.

In termini moderni, queste dimostrazioni pseudogeometriche non hanno assolutamente nulla a che fare con la teoria copernicana. Agli occhi di Bruno, esse trasformano la teoria copernicana in una grande metafora (o, come avrebbe detto piuttosto Bruno, in un emblema) della propria filosofia. Se ci rendiamo conto che Bruno stava usando la teoria copernicana semplicemente come una metafora filosofica, i suoi grossolani errori matematici e fisici diventano insignificanti. Se ci fosse bisogno di altre prove del suo totale disinteresse per la propria incomprensione dei particolari essenziali della concezione eliocentrica, potremmo ricordare la sua sconfitta ignominiosa in una discussione sull'ipotesi copernicana tenutasi a Oxford nel 1583. Richiamandosi al *De revolutionibus* in un punto della discussione, pare che Bruno fraintese un diagramma. Fatto ancor piú notevole, la sua difesa di Copernico si fondò non tanto sul *De revolutionibus* quanto piuttosto su un'opera sulla magia del neoplatonico Marsilio Ficino. L'« incidente » viene trasformato in *La cena de le ceneri*, dove Bruno appare il vincitore della disputa, pur mantenendo il proprio fraintendimento. Egli pone la Terra e la Luna sullo stesso epiciclo, invece di far ruotare l'epiciclo della Luna attorno alla Terra. Una prova ancora piú convincente della mancanza di interesse in Bruno per i particolari astronomici è un'asserzione esplicita del Nolano nel dialogo: « Io curo poco il Copernico..., et poco mi curo che voi o altri l'intendano ». E in un altro punto afferma che non si preoccupa affatto di cose come l'ottica.

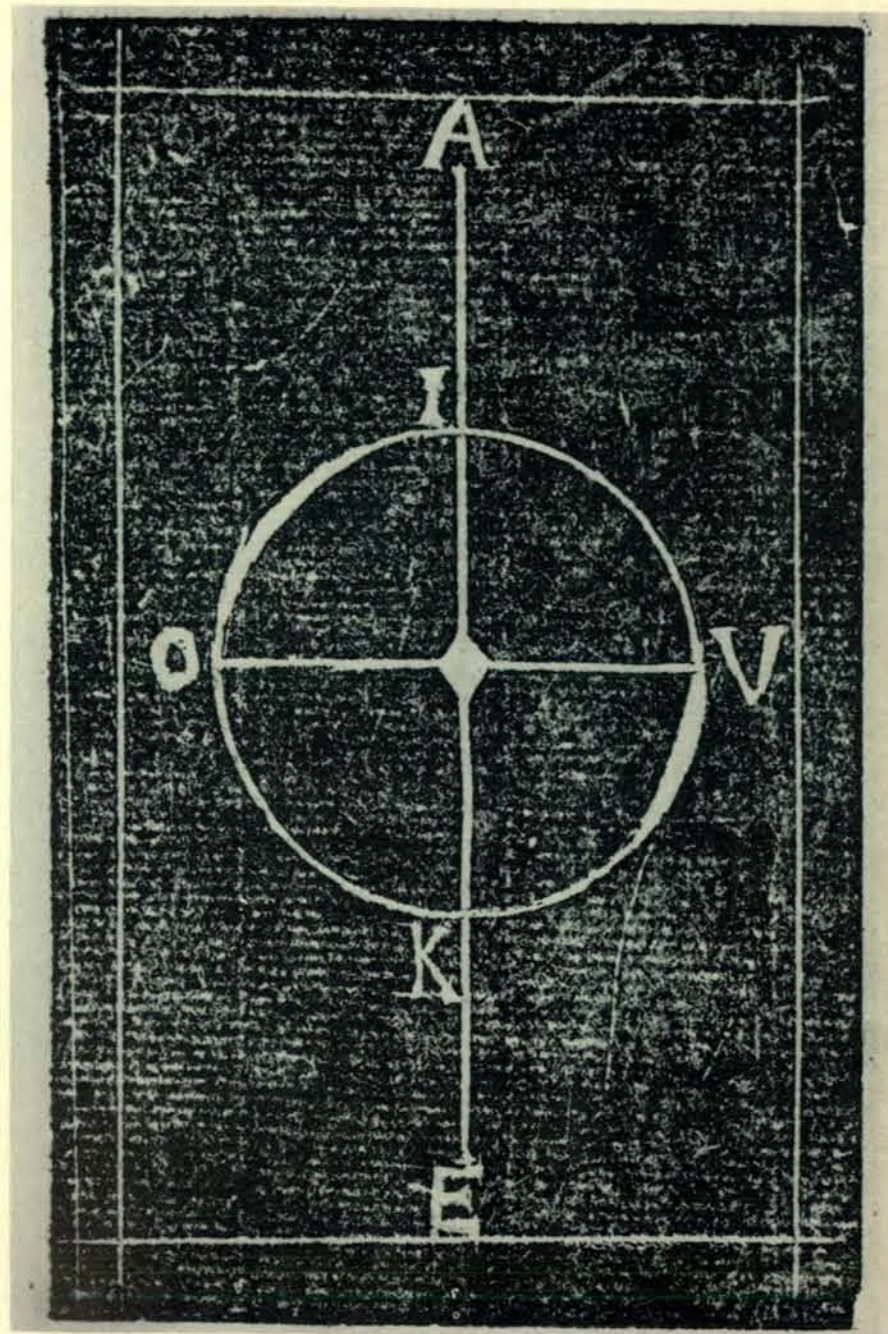
Il vero interesse del Bruno per l'eliocentrismo risiede nel fatto che esso nega che la Terra sia l'unico centro dell'universo. Prendendo di qui l'avvio egli poteva esporre la sua propria concezione secondo cui l'universo è infinito in estensione e contiene un numero infinito di mondi, ciascuno dei quali può essere considerato il centro dell'universo alla stessa stregua di qualsasi altro. La nozione bruniana di un universo privo di centro o con molti centri era derivata dal platonico quattrocentesco Nicola Cusano, che aveva parlato, in termini filosofici, di un universo di dimensioni imprecisate, il cui « centro è ovunque e la circonferenza in nessun luogo ». Di fatto questa formula comprendeva anche la definizione ermetica (e bruniana) dell'uomo. Secondo Bruno l'universo presenta, sul piano fisico, quella stessa infinità che la mente umana rivela sul piano intellettuale. L'universo è perciò una creazione adeguata di un Dio infinito che è Tutto in Tutto. Esso è anche un oggetto idoneo di contemplazione per quel ricettacolo infinito che è la mente dell'uomo. L'elemento mistico presente in tutto ciò è perfettamente chiaro. Secondo la psicologia aristotelico-neoplatonica, la mente diventa ciò che contempla. Il microcosmo-uomo si trasforma cosí nel macrocosmo-universo e questo fatto lo avvicina al Creatore.

Alla luce di queste argomentazioni emerge un'immagine di Bruno del tutto diversa da quella iniziale. Egli era un vero uomo del Rinascimento e non può essere considerato moderno sotto nessun aspetto. Bruno riflettè il modo di pensare globale del Rinascimento, nella convinzione che ogni conoscenza è interessante perché si inserisce in un tessuto di conoscenza totale. Il fine di questo tessuto globale di conoscenze, e la chiave della sua comprensione, risiedono nella posizione dell'uomo nell'universo, ossia nella relazione tra l'uomo e Dio. La conoscenza è utile nella misura in cui chiarisce questa relazione essenziale. Come ogni conoscenza, la conoscenza fisica dell'universo viene considerata una parte di una grandiosa metafora. I particolari hanno poca importanza finché non disturbano la metafora.

Benché Bruno descrivesse infine taluni aspetti centrali della teoria di Copernico, la sua descrizione dei moti della Terra differisce significativamente da quella di Copernico. Questa discrepanza è attribuibile in parte alla limitata comprensione, da parte di Bruno, dei particolari astronomici, e in parte al suo desiderio di migliorare l'accordo tra l'astronomia di Copernico e la propria filosofia. Il vero interesse di Bruno andava alla costruzione di un'immagine di grande portata, convincente ed esteticamente valida che conducesse a un'illuminazione metafisica. In questo senso non moderno Bruno faceva ricorso al copernicanesimo.

Siamo ora in grado di considerare in che modo la scienza e la filosofia di Bruno furono espresse nel suo programma di azione pratica. Il fine di Bruno era, come abbiamo osservato, la riconciliazione dell'uomo con Dio. Il XVI secolo era un'epoca di lacerazione religiosa. Dovunque i cristiani si uccidevano fra loro nel nome di Dio e questa situazione sembrava destinata a protrarsi all'infinito. La concezione teologica che sottolineava il carattere malvagio della natura umana sembrava ampiamente giustificata dall'esperienza. Benché la nuova religione protestante sostenesse di rifarsi ai principi cristiani originari, essa sembrava altrettanto impotente del cattolicesimo a produrre una moralità pubblica universale. Né il protestantesimo né il cattolicesimo apparivano in grado di mettere efficacemente freno alla corruzione e a un comportamento cinico da parte degli individui o del pubblico. Sia nei paesi cattolici sia nei paesi protestanti il mondo della cultura era dominato da una proliferazione di aridi pedanti.

Il neoplatonismo ermetico di Bruno propose una via d'uscita. Bruno credeva, come molti altri pensatori del Rinascimento, che fosse esistita un'età dell'oro in cui gli uomini dovevano essere stati fisicamente e moralmente superiori a quelli della propria epoca. Bruno e gli altri neoplatonici ermetici usavano come base della loro prospettiva filosofica e storica un insieme di libri chiamati *Corpus Hermeticum*. Il *Corpus Hermeticum* era attribuito a Ermete Trismegisto (Ermete il tre volte grandissimo), un mago-filosofo-teologo egiziano che si riteneva fosse stato pressappoco contemporaneo di Mosé.

Si riteneva che la « vera religione » rivelata a Ermete Trismegisto fosse stata tramandata da un lato dai filosofi greci pitagorici e platonici e dall'altro dai teologi ebraico-cristiani. (Quest'opinione condusse infine Bruno alla conclusione che il Cristo non fosse di origine divina ma che fosse semplicemente un mago.) Lungo questa tradizione, la verità era stata corrotta e diluita, e questo spiegava la condizione di declino in cui si trovava l'uomo del XVI secolo. Solo riportando in vita la vera religione si poteva tentare una riconci-

Una palla scagliata verso l'alto illustra secondo Bruno i quattro movimenti copernicani della Terra. Il movimento verso l'alto della palla da *A* e *B* (*non illustrato*) corrisponde alla rivoluzione annua della Terra attorno al Sole. La rotazione diurna della Terra sul suo centro, la quale dà origine a uno scambio continuo dei punti rappresentati da *I* e da *K*. Inoltre i punti 1, 2, 3, e 4 (*non indicati*) sulla circonferenza si sposteranno nei punti 5, 6, 7 e 8 (*anch'essi non indicati*). Infine, la palla si sposta obliquamente, cosicché *I* e *K* non percorrono mai una stessa linea retta e i punti *O* e *V* mutano posizione. Gli ultimi due movimenti, descritti in modo alquanto ambiguo, dovrebbero corrispondere al terzo moto (da Copernico attribuito alla Terra a causa della mancanza della nozione d'inerzia) e al quarto (la precessione). Il fatto che Bruno non abbia indicato nel diagramma i punti segnalati nel testo risponde probabilmente a un'intenzione mistica.

liazione dei cattolici e dei protestanti e una riforma della società.

In contrasto con la nozione cristiana ortodossa che l'uomo fosse caduto da uno stato di grazia a causa del peccato originale, gli ermetisti ritenevano che l'uomo fosse disceso volontariamente dal mondo immateriale della mente divina alla Terra e continuasse a partecipare della natura divina che aveva avuto prima della discesa. Di fatto, tra l'uomo nel suo aspetto divino e Dio esisteva un canale che consentiva una comunione continua; questo canale poteva essere approfondito e ampliato acquistando una sorta di conoscenza recondita che noi chiameremo occulta, la quale si fondava su talune affinità tra l'uomo e il cosmo che erano state esaminate da Nicola Cusano. In conseguenza della curiosa giustapposizione tra la fede nella magia e quella nella ragione come vie alla conoscenza di Dio, molti studiosi seri, tra cui Dee, Gilbert, Ficino e Paracelso, si dedicarono simultaneamente (e senza sentirsi colpevoli di incoerenza) a investigazioni magiche e scientifiche.

L'uomo avrebbe raggiunto la sua redenzione sia materiale sia spirituale rendendosi conto di quest'intima connessione esistente tra lui e Dio. Questa connessione mistica forniva inoltre un mezzo per agire sull'universo. Se Dio poteva usare tale connessione in un senso « discendente » per compiere taluni fatti soprannaturali, oltre a quelli naturali, l'uomo avrebbe potuto fare qualche cosa di simile agendo in senso « ascendente » mediante il ricorso alla magia. Adottando questo punto di vista, diventava indistinto il confine

L'« esperimento della nave » fu un altro degli esperimenti mentali di Bruno. Una nave procede con moto regolare lungo la corrente di un fiume. Se un uomo sull'albero della nave e un uomo sulla riva lasciano cadere un sasso ciascuno, dallo stesso punto e nello stesso tempo, la pietra lasciata cadere dalla sommità dell'albero batterà al piede dell'albero, quale sia la velocità della nave, mentre l'altro sasso cadrà verso poppa. Tra il testo di Bruno e la xilografia ci sono numerose discrepanze. Il testo sembra riferirsi a un disegno geometrico, citando punti *A*, *B*, *C* ecc. nessuno dei quali appare nell'illustrazione. Non si tratta di semplice trascuratezza, ma di un fatto legato a tradizioni iconografiche del XVI secolo. Nessuno dei due sperimentatori nominati nel testo è visibile, né lo sono i sassi da essi lasciati cadere. La nave è vista da poppa, ossia dalla posizione peggiore per chi volesse osservare la differenza nella traiettoria seguita dai due sassi. Invece che su un placido fiume, la nave è raffigurata su un mare turbolento, che rappresenta i conflitti politici e religiosi dell'epoca di Bruno. Il dio del vento rappresenta le cause spirituali sottostanti a tali conflitti. Le fiammelle che escono dai bracci del pennone rappresentano gli dei stellari Castore e Polluce, i Gemelli, che simboleggiano la tranquillità spirituale. Bruno introdusse questi simboli non solo per illustrare il testo, ma per guidare il lettore al di là di esso in un modo quasi mistico.

tra l'investigazione magica e quella scientifica, che in tempi moderni è invece assai netto. I neoplatonici non vedevano alcuna incompatibilità tra le due attività, che consideravano piuttosto complementari e reciprocamente corroborantisi.

Riesaminiamo ora gli argomenti « ottici » di Bruno alla luce della sua filosofia. È assurdo sostenere che una sorgente luminosa sufficientemente lontana possa illuminare un'intera sfera. Ma Bruno parla di una luce divina, non di una luce fisica. È evidente che le leggi dell'ottica fisica non limitano il flusso di questa luce spirituale che va da Dio all'uomo dovunque quest'ultimo possa trovarsi, in senso figurato come in senso fisico. L'affermazione chiude in sé una ricca immagine multipla che val la pena di considerare dettagliatamente sia per il suo valore poetico sia per la luce (in senso bruniano!) che proietta sul modo in cui l'uomo del Rinascimento integrava quella che chiameremo scienza nella totalità del suo pensiero.

Ci sono buone ragioni per credere che Bruno si riferisca implicitamente, soprattutto, ad argomentazioni proposte dai Padri della Chiesa Lattanzio e Agostino. Sia o no piatta la Terra, ovvero se essa è sferica, non possono esistere uomini che vivono agli antipodi (ossia dall'altra parte). Se ce ne fossero, il loro su sarebbe il nostro giú e la loro estate il nostro inverno. Inoltre la discesa di Cristo all'ultimo avvento non potrebbe esser vista da tutti. Poiché gli antichi non potevano recarsi agli antipodi, la gente che vi vivesse non potrebbe discendere da Adamo, essere stata contaminata dal peccato originale ed essere capace di redimersi. Per estensione, tali uomini non sarebbero soggetti a un papa del XVI secolo. Le conoscenze acquisite mediante le grandi esplorazioni avevano ovviamente screditato questa teoria. Bruno sta dicendo in modo allegorico che, attraverso le nuove conoscenze da lui comunicate, può essere compiuta la controparte metafisica di quelle esplorazioni. L'illuminazione divina può riconciliare i poli apparentemente opposti del cattolicesimo e del protestantesimo.

L'argomentazione del corpo opaco che separa i due corpi luminosi ha un intento simile. La differenza dottrinale principale tra cattolici e protestanti consisteva (e ancora consiste) in una discussione sull'esatta natura dell'ostia eucaristica. Secondo i cattolici l'esistenza fisica del pane e del vino viene obliterata e trasformata miracolosamente nella sostanza fisica del corpo e del sangue di Cristo. I protestanti hanno in proposito varie opinioni; quel che rimane costante è la convinzione che il pane e il vino conservino la propria esistenza fisica. Ciò che Bruno intende dire è che una semplice sostanza fisica opaca, come l'ostia eucaristica, sia essa o no transustanziata, non può (o non deve) opporsi alla riconciliazione e all'unità delle anime, cattoliche e protestanti, a opera della mente divina. In quest'immagine le anime sono rappresentate dai corpi luminosi e la luce è l'illuminazione divina, trasmessa da ciascuno dei due corpi all'altro.

Lo strano titolo del dialogo di Bruno, che non appare affatto appropriato ammettendo che l'argomento trattato sia il copernicanesimo, acquista il suo intero significato alla luce del messaggio arcano di Bruno. *La cena de le ceneri* riguarda il banchetto eucaristico, o della comunione. Il significato simbolico delle « ceneri » proviene da un mito antico citato espressamente da Bruno. Secondo il mito, un uomo che aveva offerto un sacrificio sulla sommità dell'Olimpo scrisse prima di andarsene alcune parole sulle ceneri raffreddate. Tornato l'anno successivo, trovò le ceneri intatte e le parole altrettanto leggibili di quando le aveva scritte l'anno prima. Quest'immagine è ovviamente appropriata a raffigurare la perpetua tranquillità presente in quella regione elevata e quasi spirituale. L'uomo può raggiungere tale tranquillità adottando il programma della rigenerazione spirituale di Bruno. Il mercoledí delle Ceneri è l'inizio della quaresima, un periodo che adombra tale rigenerazione; anche qui la pertinenza dell'allusione alle ceneri risulta chiara.

Se l'uomo moderno può apprezzare facilmente la ricchezza d'immagini degli scritti di Bruno, non presta fede altrettanto facilmente alla nozione che Bruno avrebbe tentato di compiere una missione pratica seria mediante un messaggio arcano come quello qui delineato. È tuttavia essenziale rendersi conto del fatto che la mentalità dei neoplatonici del XVI secolo aveva una forte inclinazione verso gli arcani e che il messaggio di Bruno era destinato all'ambiente di corte ermetico-neoplatonico della regina Elisabetta. In tale ambiente gli emblemi di Bruno e i suoi simboli ottici metafisici sarebbero risultati palesi. Considerato in questa luce poetica e metafisica, l'uso da parte di Bruno della teoria copernicana come di un geroglifico metafisico sarebbe risultato manifesto ai suoi lettori. Il ricorso a espedienti del genere non era insolito e Bruno stesso se ne valse molto sovente nei suoi diversi scritti.

Bruno descrive la teoria copernicana come un elleboro, ossia un'erba che si supponeva fosse un rimedio specifico per il trattamento dell'infermità mentale e quindi un farmaco salutare per la follia religiosa e politica del suo tempo. Possiamo ora comprendere perché gli inquisitori di Bruno non abbiano mai insistito sul suo copernicanesimo in senso letterale. Essi si rendevano chiaramente conto che la teoria scientifica di Copernico non aveva alcuna relazione con l'uso – filosofico, religioso e politico – che ne fece Bruno.

Desideriamo anche suggerire qui, senza peraltro poter sviluppare questa ipotesi, la necessità di modificare la teoria generalmente accettata sulle ragioni che condussero al processo di Galileo. Ci sono prove notevoli a favore dell'ipotesi che, in circostanze politiche simili, durante la Guerra dei Trent'anni (1618-1648), Galileo fu considerato dalle autorità ecclesiastiche una sorta di Bruno redivivo. Galileo era ben consapevole del pericolo che la sua opera potesse essere interpretata come bruniana. Egli protestò ripetutamente e con energia che i suoi fini erano limitati all'investigazione scientifica e che i suoi scritti dovevano essere interpretati in modo del tutto letterale. Le sue proteste furono vane. La condanna di Galileo nel 1633 potrebbe avere avuto come obiettivo principale (anche se male inteso) la condanna di alleanze tra cattolici e protestanti del tipo di quelle patrocinate da Bruno. Alleanze del genere erano già state consumate, con grande disappunto del papato.

Citiamo qui un solo esempio dei profondi sospetti dell'Inquisizione che gli obiettivi ultimi di Galileo fossero filosofici, teologici e politici – ossia bruniani –, e non scientifici. All'uscita del volume di Galileo *Dialogo sopra i due massimi sistemi del mondo*, il padre maestro del Sacro Palazzo, Niccolò Riccardi, rivelò un grande interesse per un simbolo stampato sul frontespizio dell'opera. Il simbolo era costituito da tre delfini disposti in cerchio, ciascuno dei quali teneva in bocca la coda del precedente. Il delfino era un simbolo astrologico comune nell'iconografia del Rinascimento. Avendo saputo da un amico di Galileo che si trattava del marchio dello stampatore e che non aveva alcun rapporto con l'opera galileiana. Il Riccardi ne fu molto sollevato e disse che l'autore poteva stare di buon animo. Purtroppo l'amico di Galileo non fu in grado di trovare sul momento un altro libro pubblicato dallo stesso stampatore che contenesse l'« impresa dei delfini », e i sospetti di Riccardi ripresero alimento. Nonostante i suoi grandi sforzi e quelli degli ecclesiastici suoi amici, Galileo non fu mai in grado di dissipare tali sospetti. Quando l'Inquisizione lo costrinse a giurare che la Terra non si muove, quel che interessava non era tanto il moto fisico quanto quel moto metaforico che Galileo *doveva* avere in mente e il cui significato doveva essere sovversivo ed eretico.

In considerazione del grande abisso esistente tra il modo di pensare di Bruno e quello di Galileo su quello che appare essere superficialmente lo stesso argomento, dobbiamo chiederci quale sia stata, se ci fu realmente, l'influenza di Bruno sulla rivoluzione scientifica. Gilbert, come Bruno, fu un filosofo ermetico, e può darsi che l'influenza di Bruno non fu soltanto di tipo diretto, personale. Pur non essendo stato un precursore né scientifico né metodologico di Galileo, crediamo che la sua influenza sulla nascita della rivoluzione scientifica sia stata profonda, per due ragioni.

Innanzitutto, Bruno aveva una suprema fiducia nel fatto che l'uomo fosse un essere divino e non semplicemente il prodotto detestabile del peccato originale, destinato a cadere sempre piú in basso in assenza di un piú o meno capriccioso intervento divino nelle cose umane. Una tale convinzione era eretica sia dal punto di vista del cattolicesimo ortodosso sia da quello del protestantesimo ortodosso, che avevano entrambi un'idea pessimistica dell'uomo naturale. La difesa, da parte di Bruno, della concezione neoplatonica fece di lui una figura di primo piano nella rinascita della fiducia dell'uomo in se stesso, una fiducia di cui non si era avuto nulla di simile dall'antichità classica in poi. Tale fiducia toccò il suo apogeo nel XVIII e nel XIX secolo, quando uomini importanti poterono credere all'avvento di una sorta di paradiso terrestre. Benché questa concezione si sia andata un po' obnubilando nel nostro secolo, costituisce ancora la base dell'azione della maggior parte dei popoli occidentali.

In secondo luogo, Bruno credeva che la via alla perfezione si identificasse con la via alla conoscenza. È vero che la concezione bruniana su quale tipo di conoscenza fosse importante era del tutto diversa da quella moderna e in se stessa non avrebbe condotto a grandi progressi in senso moderno. Nondimeno, la convinzione che l'uomo dovesse mirare alla conoscenza piuttosto che alla fede per risolvere i suoi problemi terreni fu indispensabile per l'avvento della rivoluzione scientifica.

La tecnologia della bicicletta

Questa efficiente macchina umana ha avuto una funzione di primaria importanza nello sviluppo dei cuscinetti a sfere, dei pneumatici, delle strutture tubolari, delle automobili e degli aerei

di S. S. Wilson

Siamo portati a considerare la bicicletta come una cosa scontata dimenticando la funzione importante che essa ha avuto nello sviluppo della tecnologia moderna. La bicicletta, il primo veicolo di massa per il trasporto delle persone, ha avuto una posizione di primo piano nella fase iniziale di sviluppo dell'automobile. Oltre dunque al suo considerevole impatto diretto sulla società, la bicicletta è stata anche responsabile, indirettamente, di sostanziali mutamenti sociali ed economici. Veicolo di notevole efficienza sia da un punto di vista strutturale che meccanico, la bicicletta, come mezzo di trasporto, continua a offrire svariati vantaggi sia nei paesi industrializzati che in quelli in via di sviluppo.

Quando si considera che l'impiego della ruota nei trasporti risale a piú di 5000 anni fa, sembra strano che il primo efficiente veicolo a ruote a spinta autonoma sia stato messo a punto solo da circa 100 anni. Come accade per la maggior parte delle invenzioni che fanno epoca, molti uomini e molte nazioni possono rivendicare di avervi avuto parte. Il primo « rivendicatore » legittimo sarebbe il barone von Drais de Sauerbrun di Baden-Württemberg, che tra il 1816 e il 1818 escogitò la « draisina » un « cavallo di legno per pedoni » a due ruote mosso dalla spinta dei piedi appoggiati direttamente per terra (*si veda l'illustrazione a fianco*). Il veicolo ebbe vita breve, e non fu mai seriamente considerato un mezzo di trasporto. Nel 1839 Kirkpatric Macmillan, un fabbro scozzese, riuscí a costruire un veicolo a due ruote con spinta a pedali, veicolo che fu copiato ma non ebbe mai un successo popolare (*si veda l'illustrazione in basso*).

La bicicletta di sicurezza Rover, costruita nel 1885 dall'inglese J. K. Starley, secondo l'opinione di molti, ha segnato il definitivo sviluppo della bicicletta. La Rover possedeva già la maggior parte delle principali caratteristiche della bicicletta moderna: trasmissione diretta dello sforzo alla ruota posteriore tramite catena articolata con alto rapporto di trasmissione, cuscinetti a sfere nei mozzi delle ruote, raggi di metallo disposti tangenzialmente, struttura tubolare leggera di acciaio e telaio romboidale. A differenza della maggior parte delle biciclette moderne, la Rover aveva due tubi curvi e mancava del tubo diagonale che va dalla sella all'albero delle pedivelle; anche la forcella anteriore, seppure inclinata, era diritta invece di essere curva. A meno che non sia altrimenti specificato, i veicoli delle fotografie che illustrano l'articolo sono al Museo della scienza di Londra.

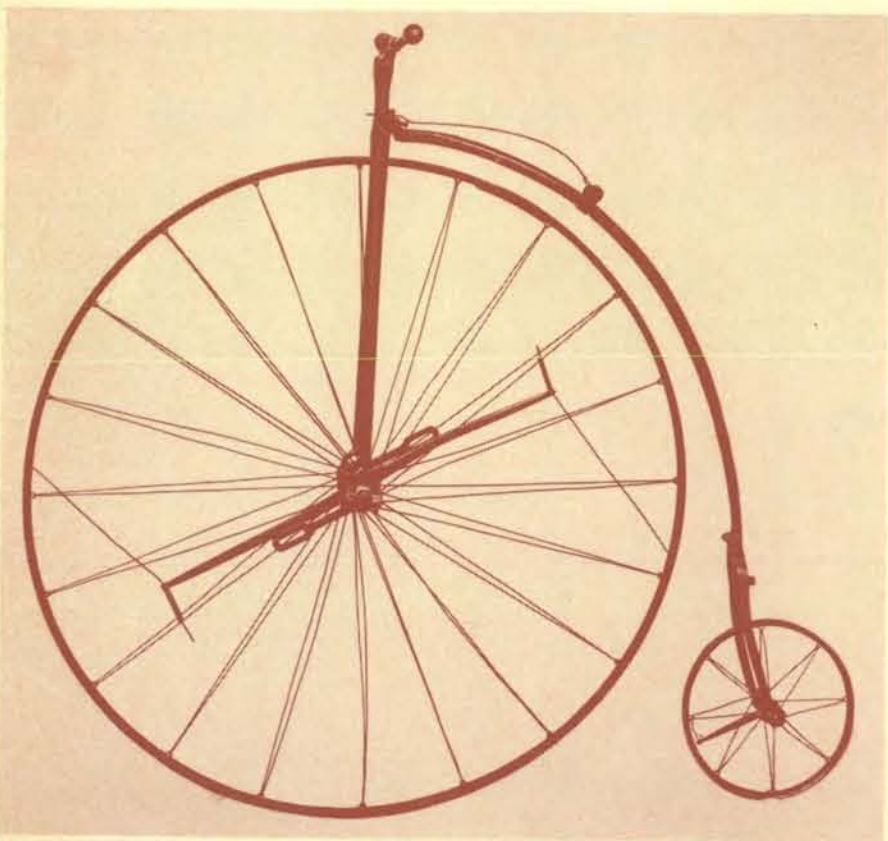
La « draisina », un « cavallo di legno per pedoni » escogitato tra il 1816 e il 1818 dal barone von Drais de Sauerbrun di Baden-Württemberg, è considerata la progenitrice della bicicletta. Il veicolo, che si muoveva grazie alla spinta diretta dei piedi del « ciclista » sul suolo, non fu però considerato un serio mezzo di trasporto. Questo particolare modello che risale al 1817 è stato fotografato nel museo di Heidelberg.

La prima macchina commercialmente importante di questa stirpe fu il velocipede francese, ideato da Pièrre e Ernest Michaux a Parigi nel 1863 (*si veda l'illustrazione in alto nella pagina seguente*). Questo veicolo, chiamato qualche volta « scuoti-ossa » aveva le pedivelle fissate direttamente al mozzo della ruota anteriore, come nel piú semplice dei tricicli per bambini. Di conseguenza il suo valore era diminuito dal fatto di avere un « rapporto di trasmissione », per usare un termine moderno, troppo basso. Ciò significava che un giro di pedali portava avanti il veicolo di una distanza uguale alla circonferenza della ruota anteriore, forse solo 300 centimetri. (In una bicicletta moderna un giro di pedali, grazie alla trasmissione a catena articolata, fa avanzare il veicolo di 500 centimetri o piú). Il solo modo per superare questo limite, pur mantenendo la semplicità della spinta diretta, era quello di adottare una ruota anteriore molto grande. Il passo successivo, quindi, nello sviluppo della bicicletta, il famoso « velocipede a ruota alta », era caratterizzato da una ruota anteriore di 180 centimetri di diametro e da una ruota posteriore con un diametro di 60 centimetri o meno.

Il velocipede a ruota alta, noto anche come « biciclo » o macchina « comune » nacque in Inghilterra a Coventry, ora centro dell'industria automobilistica britannica. La sua invenzione fu dovuta in larga parte al lavoro di una famiglia, gli Starleys. Il succedersi degli avvenimenti fu determinato dal fatto che la Coventry Sewing Machine Company aveva un rappresentante a Parigi, un certo Rowley Turner che nel 1868 portò a Coventry un velocipede Michaux. James Starley, tecnico e inventore autodidatta, era a quel tempo direttore delle officine e immediatamente vide le possibilità che offriva la produzione di biciclette. La ditta divenne presto la Coventry Machinists' Company e cominciò a prendere ordinazioni per la fabbricazione di parecchie centinaia di veicoli del tipo Michaux per il mercato parigino. Proprio allora scoppiò la guerra franco-prussiana del 1870 cosicché i veicoli furono venduti soprattutto in Inghilterra. Questo episodio portò a un periodo di sviluppo tecnico e commerciale intenso che ebbe come conseguenza non solo il perfezionamento della bicicletta nella sua forma definitiva ma anche la comparsa della motocicletta, del triciclo a motore e dell'automobile.

Prima di considerare nei particolari questo succedersi di progressi, val la pena di chiedersi perché un'invenzione apparentemente semplice come la bicicletta dovrebbe avere dato un impulso cosí grande al progredire della tecnologia. La risposta si trova sicuramente nel valore che questa macchina ha per l'uomo. Scopo della bicicletta è quello di rendere piú agevoli gli spostamenti e questo scopo è raggiunto in modo da superare nettamente l'evoluzione naturale. Se si confronta l'energia consumata nel percorrere una certa distanza da numerosi animali e macchine in rapporto al loro peso si trova che un uomo che cammina senza aiuto si comporta piuttosto bene (consumando circa 0,75 calorie per grammo al chilometro), ma non è efficiente come il cavallo, il salmone o un aereo a reazione (*si veda l'illustrazione a pag. 41*). Con l'aiuto di una bicicletta, però, il consumo energetico dell'uomo su una data distanza è ridotto grosso modo a circa un quinto (0,15 calorie per grammo al chilometro). Quindi, a parte l'aumento della velocità di un fattore di

Biciclo spinto a pedali costruito nel 1839 da Kirkpatric Macmillan, un fabbro scozzese, per suo uso e consumo. Anche se fu imitato, non ebbe mai un successo popolare.

Il velocipede francese, costruito da Pierre e Ernest Michaux a Parigi nel 1863, fu il primo veicolo commercialmente importante nella storia che precedette l'avvento della moderna bicicletta. Lo « scuoti-ossa », come veniva a volte chiamato, aveva le pedivelle che portavano i pedali fissate direttamente al mozzo della ruota anteriore e quindi era limitato nell'efficienza dal fatto di avere un rapporto di trasmissione troppo basso.

Il velocipede a ruota alta, ideato per la prima volta dalla famiglia Starley, fu costruito per ovviare al basso rapporto di trasmissione del velocipede del tipo Michaux, però conservava l'attacco diretto dei pedali al mozzo della ruota. Il velocipede a ruota alta, noto anche come « biciclo » o macchina « comune » aveva una ruota anteriore con un diametro di 180 centimetri e una ruota posteriore di soli 60 centimetri di diametro o anche meno. Questo particolare modello, chiamato Ariel, fu prodotto nel 1870 da James Starley e William Hillman. I raggi, disposti radialmente, non erano molto adatti a resistere alla forte torsione esercitata dai pedali sul mozzo. Per questo motivo, furono aggiunte due sbarre rigide supplementari, ciascuna con un suo raggio regolabile, per contribuire a trasmettere la forza di torsione dal mozzo al cerchione. La soluzione definitiva, la ruota a raggi tangenziali, fu brevettata da Starley quattro anni piú tardi.

tre o quattro, il ciclista migliora la sua efficienza salendo al primo posto tra gli animali e le macchine che si muovono.

Per giungere a questa eccellente prestazione, la bicicletta si è via via perfezionata cosí da raggiungere, da un punto di vista ergonomico, la struttura ottimale. Essa utilizza i muscoli giustí (quelli delle cosce, i piú potenti del corpo) facendo compiere loro un giusto movimento (una leggera azione rotatoria del piede) alla giusta velocità (da 60 a 80 giri al minuto). Una tale struttura deve poter trasmettere potenza in maniera efficiente (per mezzo di cuscinetti a sfere e di una catena a rulli cavi); deve rendere minima la resistenza al rotolamento (per mezzo di pneumatici) e avere un basso peso per ridurre lo sforzo del pedalare in salita.

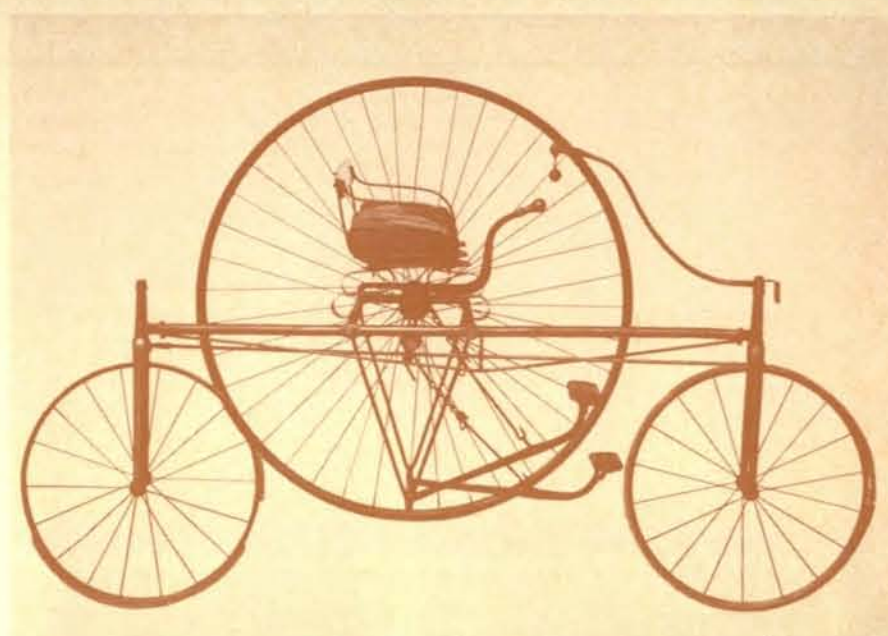
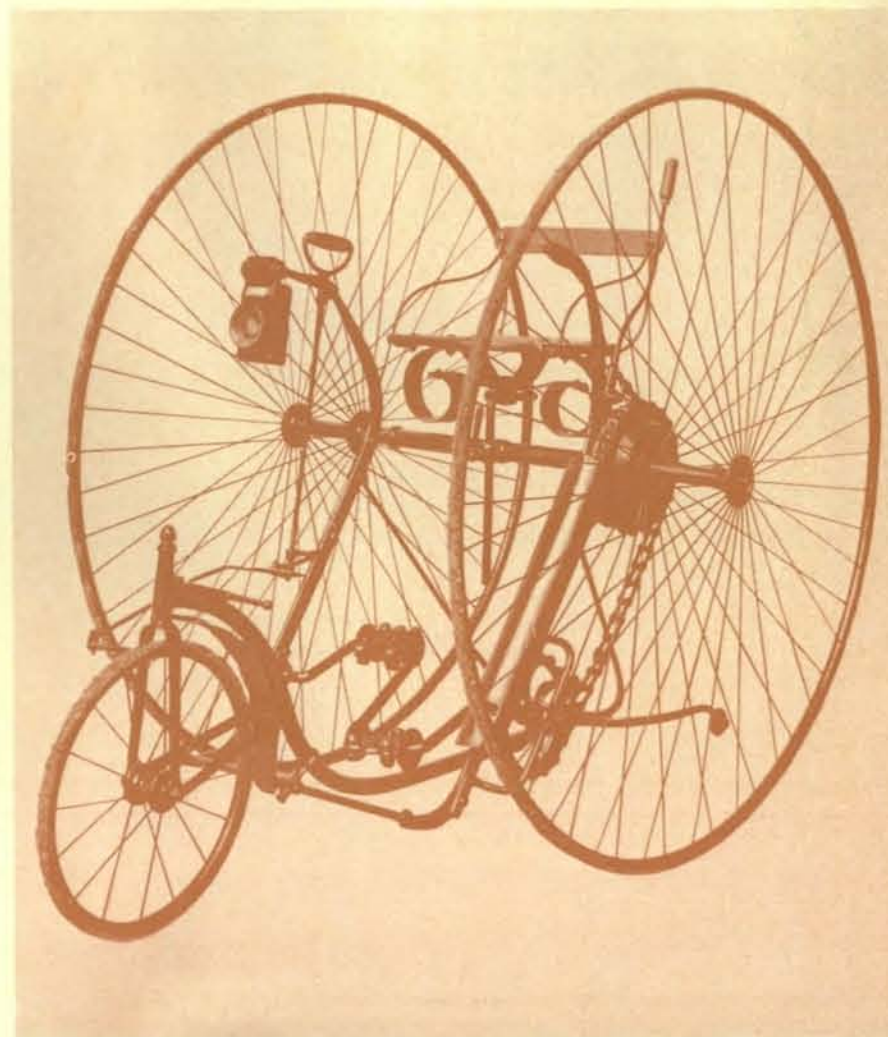
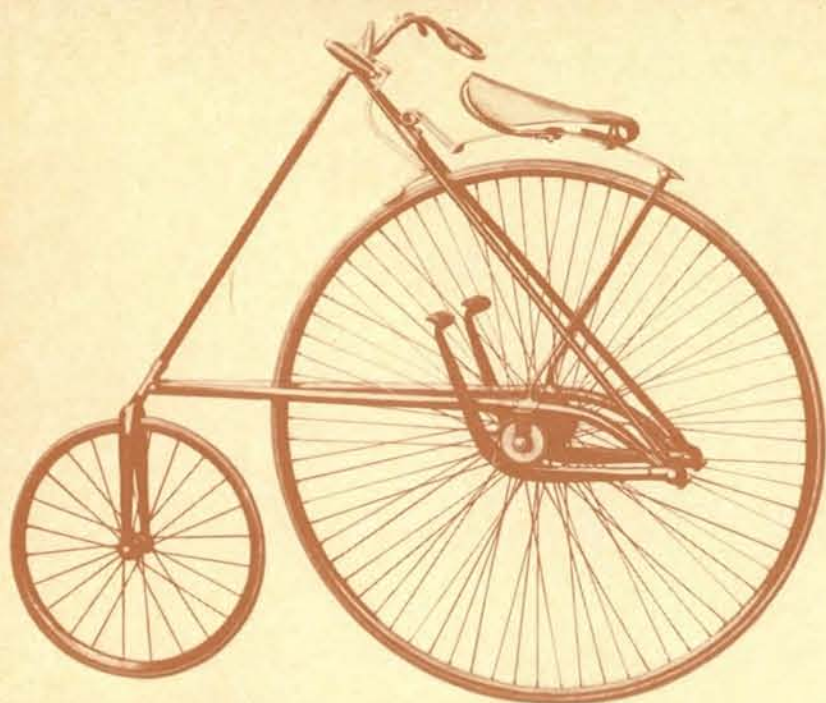
La ragione della elevata efficienza energetica dell'andare in bicicletta rispetto al camminare sembra risieda soprattutto nelle modalità di azione muscolare. Mentre una macchina compie un lavoro meccanico solo quando una forza si sposta su una distanza, i muscoli consumano energia anche quando sono in tensione ma non in movimento (compiendo lavoro a volte detto « isometrico »). Un uomo che sta in piedi si mantiene in posizione eretta per mezzo di un complesso sistema di ossa in compressione e di muscoli in tensione. Quindi il semplice stare in piedi consuma energia. Analogamente, nel compiere movimenti non contrastati da forze esterne, come avviene nell'allenamento « contro l'ombra » della boxe, si consuma energia muscolare a causa della accelerazione e decelerazione alternate delle mani e delle braccia, nonostante non si compia alcun lavoro meccanico contro un'azione esterna.

Camminando, i muscoli delle gambe non devono soltanto sostenere il resto del corpo in posizione eretta ma anche alzare e abbassare tutto il corpo cosí come accelerare e decelerare gli arti inferiori. Tutte queste azioni consumano energia senza compiere alcun lavoro esterno utile. Camminare in salita richiede che venga compiuto un ulteriore lavoro contro la forza di gravità. Oltre a queste forme di consumo energetico, si perde energia ogni volta che il piede tocca il suolo, come si può constatare dal logoramento dei pavimenti, delle scarpe e delle calze. Anche l'oscillazione delle braccia e delle gambe causa consumo e perdita di energia per riscaldamento.

Questo non avviene per il ciclista che, prima di tutto, risparmia energia stando seduto, alleggerendo cosí i muscoli delle gambe dalla funzione di sostegno e dal conseguente consumo di energia. Le uniche parti del corpo che vanno avanti e indietro sono le ginocchia e le cosce; il piede ruota leggermente a velocità costante e il resto del corpo è fermo. Perfino l'accelerazione e la decelerazione delle gambe sono ottenute in maniera conveniente, dato che vengono usati quasi esclusivamente i muscoli piú robusti; la gamba che si innalza non deve essere sollevata ma è portata in alto dalla spinta verso il basso dell'altra gamba. I muscoli dorsali devono essere usati per sostenere il busto, ma anche le braccia possono contribuirvi (nella normale posizione che una persona assume andando in bicicletta) con uno sforzo minimo delle mani e delle braccia.

Il corridore ciclista assume invece una posizione meno confortevole allo scopo di diminuire la resistenza opposta dal vento, l'elemento forse piú negativo della bicicletta rispetto alla perdita di energia. La resistenza del vento è proporzionale al quadrato della velocità del vento rispetto al ciclista. Di conseguenza se si va in bicicletta a 20 chilometri l'ora con vento contrario che soffia a 10 chilometri l'ora, la resistenza del vento sarà nove volte maggiore che se si dovesse tenere la stessa velocità con vento a favore di 10 chilometri l'ora. In pratica, come tutti sanno, si può regolare la velocità adattandola alle condizioni del vento variando il rapporto del cambio per mantenere una velocità ottimale di pedalata. Oltre alla resistenza del vento, la sola perdita significativa di energia è quella dovuta all'attrito di rotolamento, che con ruote di normale grandezza e con pneumatici adeguatamente gonfiati è molto piccolo su una superficie liscia ed è quasi indipendente dalla velocità.

Poiché, nella bicicletta, ogni parte componente deve essere dimensionata alla struttura umana, l'intera bicicletta deve essere sempre proporzionata e in scala umana. La leggerezza della costruzione, ottenuta soprattutto adottando ruote a raggi e telaio tubolare, fu auspicata non solo per il fatto che la bicicletta doveva affrontare delle salite ma anche per il desiderio di renderla facilmente sollevabile. Poiché la bicicletta richiede per la sua costruzione poco materiale e limitate risorse di energia, è scarsamente inquinante, dà un contributo positivo alla salute e causa raramente gravi incidenti, può essere considerata la piú generosa delle macchine.

Per ritornare alla storia della bicicletta, nel 1870 James Starley e William Hillman (che in seguito fondò la casa automobilistica che porta il suo nome)

Il triciclo a leva Coventry fu progettato da James Starley nel 1876 con lo scopo di aggirare le difficoltà di salire e di restare in sella sul velocipede a ruota alta.

Il triciclo Royal Salvo, anch'esso progettato da James Starley, attirò l'attenzione della regina Vittoria che ne ordinò due esemplari, decretando cosí la « rispettabilità » della nuova mania di andare in bicicletta. Questo triciclo, che incorporava uno dei primi differenziali, inventato da Starley nel 1873, fu considerato particolarmente adatto alle signore.

La bicicletta Star, costruita dalla Smith Machine Company nel 1881, fu un altro tentativo di rendere piú sicuro il velocipede. Questo veicolo aveva una piccola ruota anteriore e un sistema di leve, rulli, e cinghie per muovere la grande ruota posteriore.

progettarono l'Ariel: un elegante velocipede tutto di metallo, compresi i raggi delle ruote (*si veda l'illustrazione in basso a pagina 32*). I raggi che erano disposti radialmente e potevano essere tesi a piacere non erano molto adatti a resistere alla forza torsionale esercitata dai pedali sul mozzo. Quattro anni dopo Starley brevettò la soluzione definitiva del problema: la ruota a raggi tangenziali. In questa struttura, ora universalmente adottata, i raggi sono disposti in doppio ordine tangenzialmente alle due flange dei mozzi, formando cosí una serie di triangoli che rendono la ruota piú resistente alla torsione durante l'accelerazione o l'arresto. Il numero normale di raggi in una bicicletta moderna è di 32 per la ruota anteriore e di 40 per quella posteriore; essi sono di spessore uniforme inspessiti alle estremità per aumentarne la robustezza, pur rimanendo molto leggeri.

I principali difetti del velocipede a ruota alta consistevano nella difficoltà di salirvi e di restarvi in sella. Per ovviare a questi inconvenienti negli anni 70 del secolo scorso furono inventati i tricicli. Ancora una volta Starley fu al primo posto con il suo triciclo a leva Coventry del 1876 (*si veda l'illustrazione in alto nella pagina precedente*). Un altro importante tentativo per rendere piú sicuro il velocipede a ruota alta fu la bicicletta Star costruita dalla Smith Machine Company del New Jersey nel 1881 (*si veda l'illustrazione in alto*), questa bicicletta aveva la ruota piccola anteriormente e impiegava un sistema di leve per trasmettere il moto alla ruota grande posteriore. La bicicletta Star ebbe abbastanza successo ma il triciclo era molto piú adatto alle signore, una delle quali attirò l'attenzione della regina Vittoria che ordinò due tricicli Starley Salvo Royal, conobbe il costruttore e gli regalò un orologio con dedica. Niente avrebbe potuto essere piú efficace per determinare la rispettabilità della nuova mania di andare in bicicletta. Questo permise alle giovani signore di buona famiglia di uscire e di allontanarsi dall'atmosfera soffocante delle loro dimore vittoriane e portò nuove libertà anche nel modo di vestire, una tendenza promossa da Amelia Jenks Bloomer. Non ci si allontana troppo dalla realtà supponendo che l'andare in bicicletta, insieme alla graduale diffusione dell'istruzione abbia avuto una funzione significativa nei primi stadi del movimento femminile verso l'eguaglianza politica ed economica dei due sessi.

Sempre agli Starley si deve in quel periodo una innovazione tecnica che doveva essere di grande importanza per l'automobile. Questa innovazione ebbe origine da un incidente cui andò incontro un triciclo a due posti affiancati su cui James Starley e uno dei suoi figli, William, spingevano ciascuno una ruota motrice. A causa della maggior energia del giovane William il padre si rovesciò in un letto di ortiche. Mentre si stava rimettendo, a James venne l'idea del differenziale (in realtà una reinvenzione), che distribuisce lo sforzo egualmente tra le due ruote e, ciononostante, permette alle ruote di girare a velocità leggermente differenti nel compiere una curva. Anche William Starley fu un inventore prolifico e morí nel 1937 con 138 brevetti a suo nome, molti dei quali riguardavano la bicicletta.

Due importanti progressi compiuti nel 1877 furono l'adozione del telaio tubolare e dei cuscinetti a sfere. In nessuno dei due casi il concetto era del tutto nuovo, ma fu il loro esteso impiego nella bicicletta a portare queste tecniche a uno sfruttamento e a un uso universali. La stessa cosa doveva ripetersi piú tardi per i pneumatici e altre innovazioni.

Il tubo a pareti sottili di sezione trasversale circolare è l'elemento strutturale piú efficiente; può resistere a tensione, compressione, flessione, torsione o alla combinazione degli sforzi esercitati sul telaio di un veicolo. Nonostante che per quanto riguarda la flessione in un particolare piano un profilato con sezione a *I* possa essere piú efficiente, se il carico di flessione può essere applicato in ogni piano, allora è preferibile un tubo sottile. È per questa ragione che i tubi sono usati in elementi di sostegno o di compressione in cui possa verificarsi un cedimento per instabilità elastica o per inarcamento. Per quanto riguarda la torsione non vi è sezione migliore di quella circolare, e per questo motivo l'albero principale di trasmissione delle automobili è costituito da un tubo. Un esempio eccellente delle proprietà di un tubo è rappresentato dalla canna di bambú largamente impiegata in Estremo Oriente per edifici, ponti, impalcature e cosí via. In realtà con il bambú si fanno anche le biciclette.

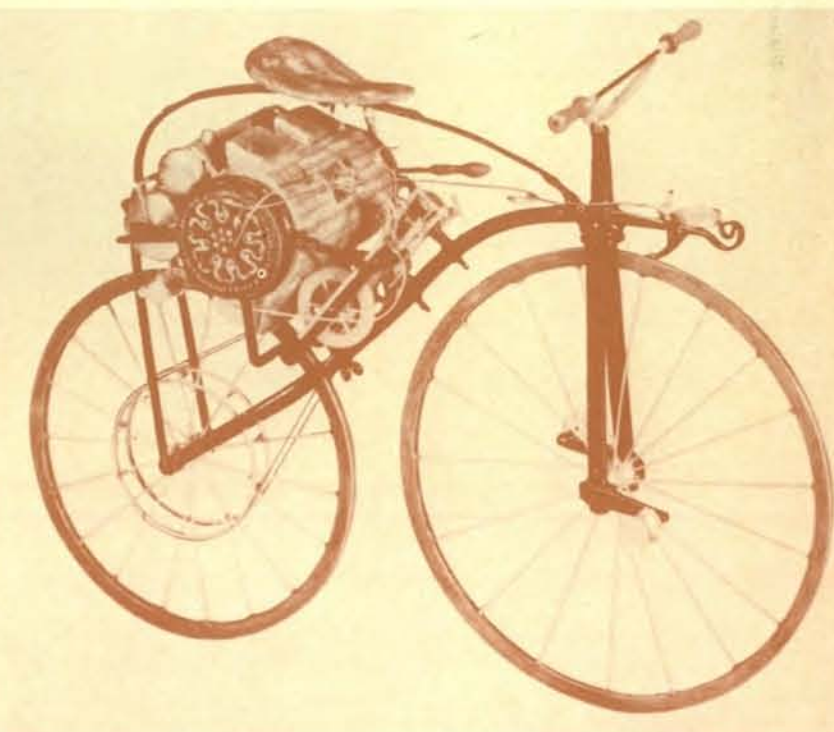
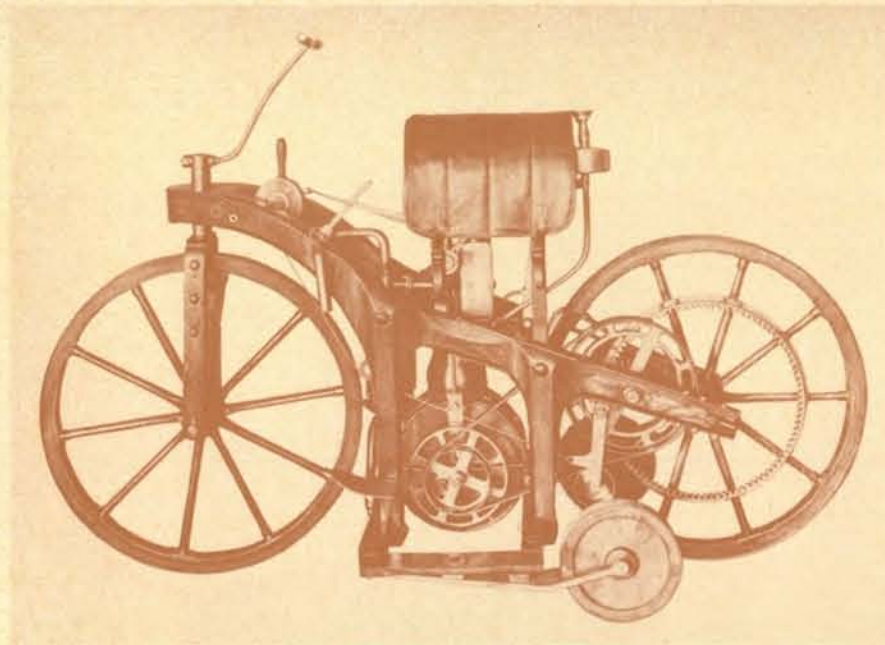
Perché si abbiano sforzi minimi, una struttura dovrebbe essere tale da subire tensione o compressione, non flessione o torsione. Questo è il principio su cui si basano i telai tridimensionali usati nelle travature dei ponti, nelle gru e nelle macchine da corsa. Tale costruzione non è pratica per la bicicletta e cosí si è giunti al compromesso del classico telaio bidimensionale a rombo. In questo telaio gli sforzi principali sono sostenuti direttamente, anche se vi sono sforzi di flessione nella forcella anteriore e sforzi di torsione nell'intero telaio quando il ciclista esercita la pressione prima su un pedale poi sull'altro. Su un velocipede a ruota alta il ciclista avverte queste forze attraverso il manubrio.

Un'alternativa al telaio romboidale apparsa per la prima volta nel 1886 e recentemente riapparsa è il telaio incrociato. Consiste di un tubo principale che va dalla sommità del manubrio sopra la ruota anteriore direttamente all'asse della ruota posteriore, incrociato da un secondo tubo che va dalla sella all'albero delle pedivelle. Si tratta di una disposizione semplice, ma si basa interamente sulla forza e sulla rigidità del tubo principale, a meno che non vengano aggiunti altri elementi per ottenere una parziale triangolazione del telaio.

A questo proposito, si ottiene un miglioramento aumentando l'area della sezione trasversale del tubo principale, raggiungendo cosí tutti i vantaggi della costruzione a tubi sottili. Questo principio, adottato nella maggior parte delle fusoliere degli aerei fin dagli anni trenta, è indicato come struttura a monoscocca o a fasciame irrigidito. Alcuni ciclomotori di recente produzione sono andati oltre e incorporano nel largo tubo principale il serbatoio del carburante, un sistema adottato anche nella costruzione di alcune macchine da corsa che hanno un serbatoio tubolare su ciascun lato del posto di guida. Nelle biciclette il vantaggio del telaio incrociato è rappresentato dal fatto che è ugualmente adatto sia per le biciclette da uomo che da donna.

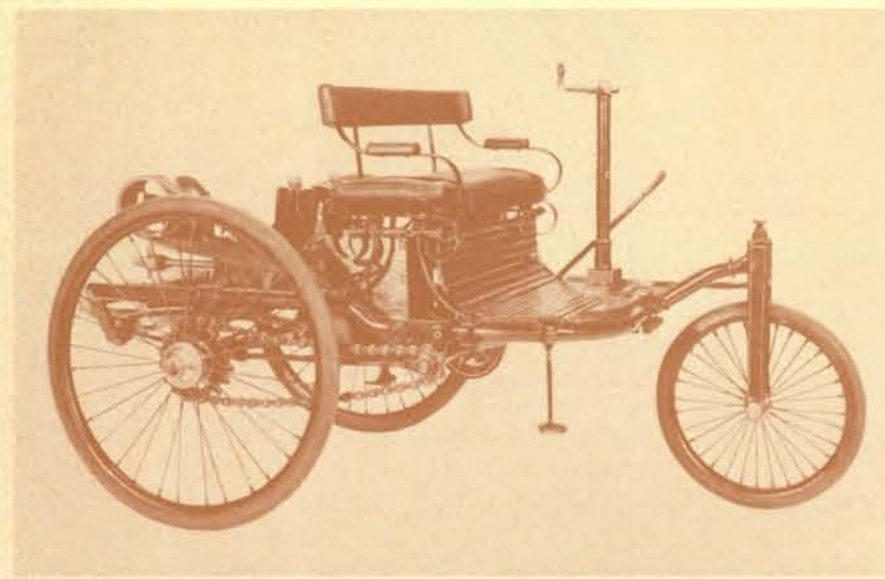
Almeno uno dei primi esemplari di telaio completamente triangolato ebbe successo: il Dursley-Pedersen, che era costruito con tubi di piccolo diametro a coppie e pesava soltanto 10 chilogrammi e mezzo (*si veda l'illustrazione in alto*). Si riuscí ben presto a ottenere biciclette poco pesanti specialmente quelle da corsa. Per esempio una comune Rudge del 1884 pesava solo 9,7 chilogrammi. Per fare un confronto, una bicicletta attuale da corsa pesa circa 9 chilogrammi e una tipica bicicletta da turismo circa 13. Normalmente si usano ancora tubi di acciaio, ma si usano anche leghe leggere, titanio e perfino plastica rinforzata con fibre di carbonio. Per formare un telaio i tubi di acciaio, generalmente al cromo-molibdeno o al manganese-molibdeno, vengono fatti combaciare e poi saldati entro manicotti di acciaio. Si ottiene cosí una struttura capace di portare un carico dieci volte superiore al suo peso, un valore non raggiunto da nessun ponte, automobile o aereo.

L'uso di un sistema a rotolamento per ridurre l'attrito trova forma definitiva nei cuscinetti a sfere o a rulli. I cuscinetti principali di una bicicletta moderna hanno almeno 12 file di sfere che rotolano tutte tra un anello conico interno e una coppa esterna. Tutte le superfici dei cuscinetti sono di acciaio rinforzato per avere una maggiore resistenza alla deformazione e al logoramento. Se questi

Il telaio completamente triangolato era la caratteristica peculiare della bicicletta Dursley-Pedersen, che ottenne un discreto successo in Inghilterra verso il 1890. Questo veicolo adottava tubi di piccolo diametro in coppie ed era notevolmente leggero.

La bicicletta a vapore fu costruita nel 1869 da Pierre Michaux utilizzando un motore Perreaux. Nonostante che i precedenti veicoli stradali a vapore si fossero dimostrati troppo pesanti e ingombranti, la tecnologia della bicicletta sembrò offrire allora nuove possibilità che però scomparvero con l'avvento del motore a combustione interna.

La bicicletta a motore Daimler, la progenitrice della moderna motocicletta, fu progettata e costruita da Gottlieb Daimler nel 1885. Il veicolo era dotato di un motore a combustione interna a un cilindro. Le due ruote ausiliarie piú piccole si ritraevano a macchina avviata. Il veicolo Daimler originale, mostrato in questa fotografia del Bettmann Archive, è ora nella ricostruita officina Daimler a Bad Cannstatt vicino a Stoccarda.

cuscinetti sono periodicamente lubrificati e adeguatamente istallati, hanno una durata sorprendentemente lunga. Perfino nel caso che vengano trascurati, continuano a funzionare a lungo e, se necessario, i vari componenti possono essere sostituiti.

Nel 1879 Harry J. Lawson fece compiere alla bicicletta un altro importante progresso: l'adozione della trasmissione a catena articolata. L'anno successivo Hans Renold costruí la forma definitiva della catena da bicicletta, la catena a rulli e boccole che unisce i vantaggi di lunga durata, efficienza e basso peso. A prima vista la struttura sembra avere scarsa ingegnosità, ma a una osservazione piú approfondita si vede esattamente come siano significative le sue varie caratteristiche (*si veda l'illustrazione a pagina 40*). La progenitrice di questa era la catena a pioli o a perni, in cui questi poggiano direttamente sui denti della ruota dentata e le piastrine di collegamento girano intorno ai perni a ciascuna estremità di questi. In questo tipo di catena c'è un esagerato logoramento e attrito sia nei denti sia nei fori delle piastrine. Un miglioramento messo a punto da James Slater nel 1864 fu la catena a rulli in cui l'attrito e il logoramento dei denti della ruota era ridotto, ma il logoramento delle piastrine sui perni era ancora troppo grande. Il progetto di Renold, con l'aggiunta ai rulli di boccole che distribuivano il carico su tutta la lunghezza del perno, fece superare questo ultimo ostacolo e diede origine alla produzione industriale delle catene di precisione. La catena a rulli e boccole si estese dalle biciclette alle macchine per l'industria tessile e ad altre applicazioni di trasmissione di energia in competizione con la « catena silenziosa » di Morse. La catena a rulli e boccole rimpiazzò la trasmissione a cinghia nelle motociclette e serví come organo di trasmissione principale per le ruote posteriori delle automobili finché non venne sostituita dalla trasmissione ad albero.

La soluzione soddisfacente del problema di una trasmissione efficiente in grado di trasmettere il moto con il rapporto desiderato rese possibile lo sviluppo finale della bicicletta nella sua forma moderna. Quest'ultimo passo fu compiuto soprattutto da J. K. Starley, un nipote di James Starley. James, onorato a Coventry come « il padre dell'industria della bicicletta », aveva formato nel 1878 con William Sutton una società per la produzione di tricicli. Nel 1885, però J. K. Starley produsse indipendentemente la famosa bicicletta di sicurezza Rover (*si veda l'illustrazione a pagina 30*) che può essere considerata il punto di arrivo della bicicletta per quanto riguarda la forma. Da quella forma la bicicletta non si è piú allontanata, nonostante un recente tentativo di adottare un telaio elastico accoppiato a ruote piccole. La bicicletta Rover aveva un telaio romboidale che incorporava due tubi ricurvi ed era priva del tubo diagonale che unisce la sella all'asse delle pedivelle. La forcella anteriore era diritta anche se inclinata. L'inclinazione era necessaria per consentire al manubrio un'azione autocentrante. La successiva adozione di forcelle anteriori curve in modo che la linea passante per il perno del manubrio incontri il suolo nel punto di contatto di

Anche il triciclo a motore Benz, il precursore della moderna automobile, fece la sua comparsa nel 1885. Questo primo tentativo di Carl Benz incorporava parti importanti quali l'accensione elettrica, un'efficace controllo della farfalla di alimentazione, valvole meccaniche, sterzo orizzontale e perfino un sedile confortevolmente imbottito. In prove compiute durante il 1886 il veicolo fu perfezionato perché mantenesse i 15 chilometri l'ora. Il veicolo Benz originale è ora al Deutsches Museum di Monaco; la fotografia mostra un duplicato dell'originale che si trova al Daimler-Benz Museum di Stoccarda.

CORONA PLANETARIA
PIGNONE PLANETARIO
MOZZO
GABBIA PLANETARIA
INNESTO SCORREVOLE
ASSE
PIGNONE SOLARE
DENTE D'ARRESTO
INNESTO SCORREVOLE
ASSE
PIGNONE PLANETARIO
MOZZO
RUOTA DENTATA
DENTE D'ARRESTO
DENTE D'ARRESTO
CUSCINETTI A SFERE
ELEMENTO MOTORE
CATENA ARTICOLATA
MOLLA DI FUNZIONAMENTO
DENTE D'ARRESTO
DENTE D'ARRESTO
ANELLO D'ARRESTO
ANELLO D'ARRESTO
PIGNONE SOLARE
CORONA PLANETARIA
GABBIA PLANETARIA

Il cambio nel mozzo Sturmey-Archer comunemente usato nelle biciclette inglesi da turismo è illustrato sia in sezione trasversale (*a sinistra*) sia in sezione longitudinale (*a destra*). In entrambi i casi gli elementi di trasmissione sono in colore scuro e gli elementi che subiscono la trasmissione in colore chiaro. In questo sistema a tre velocità è impiegato un unico treno epicicloidale di ingranaggi in modo tale che quando la bicicletta si trova nella marcia alta (*a*), la gabbia è mossa dalla ruota dentata e la ruota dalla corona planetaria. Quando la bicicletta è nella marcia media (*b*), la trasmissione è diretta alla ruota. Quando è nella marcia bassa (*c*), la ruota dentata muove la corona e la gabbia muove la ruota a una velocità ridotta. Se il pignone ha lo stesso numero di denti dei satelliti, e la corona planetaria ha tre volte questo numero, allora la marcia alta avrà un rapporto moltiplicante di 4/3 confrontato con la trasmissione diretta e la marcia bassa un rapporto demoltiplicante di 3/4.

questo con il pneumatico ha come conseguenza la riduzione dello sforzo di sterzata perché le forze laterali non tendono a ritornare al manubrio.

La comparsa della bicicletta di sicurezza Rover diede inizio al boom delle biciclette che si affermano rapidamente come quotidiano mezzo di trasporto, veicolo sportivo e mezzo per turismo su lunghe distanze. Per Coventry e per altre città dell'Inghilterra centrale vi fu un'espansione commerciale che durò fino al 1898. Poi vi fu un disastroso crollo dovuto soprattutto alle manovre finanziarie di un certo Terah Hooley. La bicicletta si diffuse in tutto il mondo e fu adattata alle necessità locali dando origine a veicoli come i risciò dell'Estremo Oriente. Tra i primi produttori di biciclette negli Stati Uniti vi furono i fratelli Duryea i costruttori della prima automobile americana nel 1893. Nel 1899 vi erano negli USA 312 fabbriche che producevano globalmente un milione di biciclette all'anno.

Prima di quest'epoca, però, vi fu ancora una conquista di grande importanza: il pneumatico. Questo elemento era stato effettivamente brevettato già nel lontano 1845 da R. W. Thomson, un ingegnere civile scozzese, allo scopo di ridurre lo sforzo per tirare le carrozze a cavalli. Non si affermò fino a quando non fu reinventato nel 1888 da John Boyd Dunlop, un medico veterinario scozzese che lavorava a Belfast. Questa volta il successo fu rapido a causa della enorme popolarità della bicicletta e all'ovvia superiorità di conforto e di efficienza del pneumatico rispetto alla gomma piena. Si aggiunsero presto ulteriori innovazioni. Charles Kingston Welch di Londra produsse nel 1890 il pneumatico con i cerchietti di filo metallico e quasi contemporaneamente William Erskine Bartlett negli USA produsse il pneumatico coi talloni. Poi David Mosely di Manchester brevettò il pneumatico a tortiglia che consisteva in strati di fili di ordito ritorti senza trama in modo da evitare un esagerato attrito interno. Perfino il pneumatico senza camera d'aria, che non è stato usato per le automobili fino agli anni 40, fu inventato per le biciclette negli ultimi anni 90 del secolo scorso. E. S. Tompkins, un esperto di pneumatici, ha scritto: « Questa ricchezza di invenzioni nei primissimi anni sorge dal fatto che gli entusiasti della bicicletta erano giovani che usavano un metodo giovane e innovatore nel progettare i pneumatici per le loro macchine ».

Un'altra conseguenza della popolarità della bicicletta fu la richiesta di

Le biciclette da record

Sul granito del monumento al ciclista che il 15 luglio verrà inaugurato sulla cima del Ghisallo, di fronte alla chiesetta dedicata alla Madonna protettrice dei corridori, ci sarà una scritta che comincia così: « ... poi Dio creò la bicicletta, perché l'uomo ne facesse strumento di fatica e d'esaltazione nell'arduo itinerario della vita... ». Che le biciclette d'oggi possano sembrare frutto d'ispirazione soprannaturale è certo, tanto sono belle, leggere, quasi impalpabili eppure tecnicamente perfette; che l'estro inventivo di appassionati artigiani abbiano contribuito a renderla così, è anche questa cosa certa. Sono settemila, uno più uno meno, i pezzi indispensabili per costruire una bicicletta e c'è modo e modo di metterli assieme.

Quella della bicicletta e del ciclismo è storia che comincia nel secolo scorso. Il mezzo meccanico nasce come audace e anche un poco snobbistica alternativa al cavallo. Progredirà di pari passo con la motorizzazione, in epoca pionieristica. Dilagherà, minaccerà di morire per soffocamento, s'imporrà di nuovo, prepotentemente, allorché – e sono i giorni in cui viviamo – la società del benessere e del malessere tenterà la fuga di fronte all'incalzare minaccioso dei tubi di scappamento, dello smog, dell'inquinamento.

Hanno cominciato i nobili a salire in bicicletta, quando le strade erano ancora in terra battuta oppure lastricate con quell'orrido selciato che faceva irritare i cavalli e sobbalzare le carrozze. Poi la mania prese quanti s'erano accontentati di guardare. Un uomo in bicicletta: ecco fatto un ciclista. Dieci, quindici uomini in bicicletta, si fissa un punto chiamato partenza, un altro chiamato arrivo, si dà un motivo per rendere valida la sfida e chi arriva primo ha vinto: ecco fatta una corsa. Il ciclismo è nato così, in Piazza d'Armi, ai bordi della città, sulle strade di campagna.

Venivano dall'Inghilterra le prime biciclette. Una « Raleigh » costava sessanta lire quando Edoardo Bianchi costruiva biciclette a 130 lire. Nacque nel 1895 la Bianchi; la Legnano – sua eterna rivale – cominciò invece nel 1906. Con loro, sorgevano la Velox, la Maino, l'Olympia, la Dei, la Frera, la Lygie, la Taurus, l'Atala, la Torpado, la Ganna. Le fabbriche di biciclette erano poco più d'una bottega. La forgia era lo strumento di base, oltre all'acciaio e anche alla gomma, visto che da poco lo scozzese J. B. Dunlop aveva inventato la gomma pneumatica che successivamente i francesi fratelli Michelin avevano reso smontabile, ma in maniera tanto complicata da convincere Giovan Battista Pirelli a cercare e a trovare qualcosa di meglio: dei pneumatici veramente smontabili in pochi minuti (ora occorrono pochissimi secondi per sostituire l'intera ruota, in corsa) grazie a certi « talloni » incorporati che tenevano la gomma fissata al cerchione mediante la pressione interna del pneumatico.

Poteva accadere di tutto nelle gare su strada dell'epoca, visto che si disputavano sotto scarso controllo, su carreggiate infami, su chilometraggi disumani. Nel 1891, il francese Carlo Terront vinceva la prima Parigi-Brest-Parigi: impiegava 61 ore e mezza a percorrere i 1208 chilometri del tracciato, e una settimana per rimettersi in salute. Nel 1896, tale Fischer s'imponeva nella prima Parigi-Roubaix, mentre Gerbi, detto « il diavolo rosso » per il colore fiammeggiante della maglia che indossava, s'aggiudicava la prima Milano-Torino (1903), il primo Giro di Lombardia (1906), il primo Giro del Piemonte (1906). Nel 1907, il francese Petit Breton si prendeva la prima edizione della Milano-Sanremo, gara improvvisata da un gruppo di fanatici sostenitori del ciclismo dopo che una Milano-Sanremo automobilistica era fallita, in quanto nessuna macchina era riuscita ad arrivare sino al traguardo.

Con la bicicletta si poteva andare dappertutto, il mondo assumeva una altra proporzione, diventava tutto da vedere: questo assicuravano produttori e venditori. Per convincere tutti, venivano organizzate gare in più giorni: il Giro di Francia nel 1903, mentre da noi, nel 1907, il Giro della Sicilia che poi sarebbe sparito, precedeva il varo del Giro d'Italia che avveniva nel 1909.

Dall'inizio del secolo, la bicicletta si può dire sia rimasta la stessa. Nella concezione di base, tale era nel 1900, tale è oggi. Ha trovato leghe nuove, sempre più leggere. Ha voluto andare sempre più veloce e questa smania l'ha soddisfatta con la scoperta del cambio di velocità (oggi, con il doppio pedaliere e i sei pignoni montati sul mozzo della ruota posteriore, è possibile affrontare con dodici « rapporti » diversi qualsiasi percorso e qualsiasi velocità). Ha sostituito la pompa a pressione con tubetti in plastica contenenti aria compressa. Ha alleggerito la catena di trasmissione, le ruote, le gomme. Ha migliorato e reso più sicuri i freni. Però, la geometria del telaio è la stessa d'allora.

Nel 1893, sulla pista in cemento di Parigi, il francese Henry Desgrange, uomo di sport e di teatro, stabiliva il primo record ufficiale dell'ora percorrendo la distanza di chilometri 35,325. Nel 1905, sulla stessa pista, il francese Petit Breton percorreva nello stesso tempo – che doveva poi diventare quello del primato che il ciclismo considera più prestigioso – la bellezza di 41,170 chilometri. Il 25 ottobre 1972, sulla pista di Città del Messico, Eddy Merckx ha portato il record a 49,43195 chilometri.

Merckx ha lasciato Desgrange a quattordici chilometri e 107 metri. In ottant'anni, tanti quanto ne vanno da un'impresa all'altra, l'uomo è notevolmente cambiato. Ha scoperto la dietetica, la medicina sportiva, la chimica, i materiali sintetici, le leghe superleggere. Anche la bicicletta, nei dettagli, ha sentito questa trasformazione.

Quella usata da Merckx, pesava appena 5,750 chilogrammi. L'ha costruita Ernesto Colnago, ex corridore, ora artigiano a Cambiago, alla periferia di Milano. Un mese e mezzo di studi e di esperimenti, gli hanno permesso di « rubare » un chilo e 150 grammi a una normale bicicletta da pista. Le caratteristiche del modello-record sono queste: telaio costruito con tubi d'acciaio dello spessore di 4/10 di millimetro e raccordi di 7/10; « gioco » dello sterzo in nylon; tubo reggi-manubrio in titanio, con ridottissimo tenditore; manubrio in alluminio, alleggerito da numerosi fori; cerchi delle ruote in duralluminio; raggi in acciaio speciale, alleggeriti (ruota anteriore 18 raggi non incrociati, ruota posteriore 32 raggi legati a tre a tre); testa dei raggi in avional; mozzi in dural con perni al titanio; tubolari in seta di 90 grammi; pedaliere a cinque braccia alleggerite mediante fresaggio; pedali senza copripolvere montati su perni al titanio, fermapiedi in dural; pedivelle alleggerite mediante fresaggio; asse del pedaliere in dural; catena a rulli alleggerita con 200 fori nelle « maglie » di congiunzione; pignone posteriore in ergal; dadi di bloccaggio delle ruote in avional; tubo reggisella in ergal.

Stupendamente collaudata, questa bicicletta resterà un prototipo: è venuta a costare non meno di mezzo milione, contro le 150-160 000 d'una normale bicicletta da pista.

La storia del ciclismo avverte a questo punto che la bicicletta costruita per Merckx non è la più leggera esistente. Sono in avanzata fase sperimentale telai costruiti in materiale sintetico, ma non di ciò si vuole parlare, bensì di una bicicletta nata nel 1948, a guerra da poco conclusa, quando l'industria che non era stata colpita a morte stentava a sopravvivere. Ebbene, in quell'epoca il capo-meccanico della Legnano, il milanese Ugo Bianchi, costruiva una bicicletta da pista che pesava 3,900 chilogrammi.

Questa macchina non poté mai essere usata in competizione, tant'era fragile e vibrante, però sotto la spinta dello stesso suo artefice che a quei tempi pesava sugli ottanta chili, ha girato per una mezz'ora, a trenta all'ora, sulla pista del Vigorelli. È ancora viva e luccicante. È stata esposta in ogni parte del globo. Telaio in acciaio specialissimo di 2/10 di millimetro di spessore, senza raccordi; mozzi da pista con l'asse forato e i coni alleggeriti con fori a monte e a valle del punto di rotolamento delle sfere; pedaliere e pedivelle costruite ricavandole, al tornio, da massello d'electron; perno e calotte del movimento centrale ricavate anch'esse al tornio, da una barra di lega leggera; catena a blocchi con maglie di congiunzione in lega al magnesio; tutti i movimenti montati su sfere alleggeriti (metà delle sfere tolte, distanziando fra di loro le rimanenti con sottilissime gabbiette in lamierino d'ottone); tubolari di 90 grammi; cerchi in alluminio a 28 raggi; sella, pedali, manubrio e fermapiedi di serie.

Questi due « prototipi » costituiscono un punto d'arrivo; ma l'epoca della missilistica in cui viviamo potrebbe trasformarli in un punto di partenza. Chissà. La bicicletta, forse, non ha ancora finito di stupire, di dare gioia. L'anno scorso, in Italia, ne sono state costruite due milioni: produzione-record. (*F.A.*)

Particolare della bicicletta con la quale Eddy Merckx ha portato il record dell'ora a 49,43195 chilometri. Allo scopo di ottenere la massima leggerezza Ernesto Colnago, l'artigiano che l'ha costruita, ha persino alleggerito con fori le maglie della catena e fresato a mano il pedaliere praticandovi delle piccole nicchie visibili nella foto.

strade migliori. In Gran Bretagna nel 1878 fu fondato il Cyclists' Touring Club e nel 1886 la Roads Improvement Association; queste iniziative contribuirono molto al miglioramento delle strade. Un movimento analogo fondato negli Stati Uniti dal colonnello Albert A. Pope di Boston, un pioniere nella costruzione di biciclette, ebbe altrettanta influenza. La bicicletta spianò letteralmente la strada all'automobile.

In che modo la bicicletta portò direttamente all'automobile è descritto dall'inventore Hiram P. Maxim, jr. in *Horseless Carriage Days* (pubblicato nel 1937): « La ragione per cui non abbiamo costruito veicoli stradali meccanici prima di allora [1890] consiste secondo me nel fatto che la bicicletta non era largamente diffusa e non aveva attirato l'attenzione degli uomini sui vantaggi del viaggiare su lunghe distanze e per strade comuni in maniera indipendente. Pensavamo che la ferrovia fosse sufficientemente buona. La bicicletta creò una nuova esigenza che la capacità della ferrovia non era in grado di soddisfare. Poi accadde che la bicicletta non poté soddisfare le esigenze che aveva creato. Era necessario un veicolo a propulsione meccanica invece di uno spinto con i piedi e ora sappiamo che l'automobile era la risposta ».

Guardando indietro dopo altri 36 anni, si è tentati di chiedersi se l'automobile è realmente una risposta tanto buona come una volta sembrava. I veicoli stradali a vapore erano stati sperimentati precedentemente, ma non erano riusciti ad affermarsi perché erano pesanti e ingombranti. La tecnologia delle biciclette leggere sembrava offrire nuove possibilità, e una motocicletta a vapore fu effettivamente costruita in Francia nel 1869 da Pièrre Michaux, usando un motore Perreaux (*si veda la illustrazione in basso a pagina 35*). Il motore a combustione interna fu applicato con successo nel 1885 da Gottlieb Daimler a un velocipede di tipo Michaux (*si veda l'illustrazione in alto a pagina 36*), e da Carl Benz a un triciclo leggero (*si veda l'illustrazione in basso a pagina 36*). Da allora in poi il progresso fu rapido, particolarmente in Francia e in seguito negli Stati Uniti.

Molti dei primi costruttori di automobili cominciarono come costruttori di biciclette, compreso Hillman, il colonnello Pope, R. E. Olds, Henry M. Leland e William Morris (piú tardi Lord Nuffield). Nel frattempo a Coventry la ditta di Starley e Sutton divenne la famosa Rover Company; produsse nel 1888 un triciclo elettrico, una motocicletta nel 1902 e un'automobile nel 1904. Sia la Morris che la Rover fanno ora parte della British Leyland Motor Corporation la piú grande casa automobilistica britannica. La prima automobile di Henry Ford usava ruote e catene da bicicletta come altri veicoli dei primi tempi come quelli tipo « buggy » (veicoli leggeri a un posto a due o a quattro ruote). I fratelli Wright erano costruttori di biciclette e le prime macchine volanti traevano un vantaggio considerevole dalla leggerezza e da altre caratteristiche strutturali che avevano ottenuto tanto successo nella bicicletta.

Forse la piú interessante applicazione della tecnologia della bicicletta al volo è l'aereo a spinta umana. Vi furono tentativi per produrre aerei di questo tipo in Germania e in Italia prima della seconda guerra mondiale, e recentemente ha suscitato molto interesse in Inghilterra l'offerta di un premio di 10 000 sterline da parte di Henry Kramer per il primo volo a spinta umana su un percorso a otto attorno a due pali distanti tra loro circa 800 metri. Si sono avuti parecchi successi nel volo in linea retta su distanze fino a circa 900 metri; il volo migliore su questa distanza è stato quello fatto dal tenente di aviazione John Potter della Royal Air Force su un monoplano progettato da Christopher Roper (*si veda l'illustrazione a pagina 42*). Questo aereo, con un'apertura alare di 25 metri e un peso di soli 66 chilogrammi, mostra chiaramente quanto si debba alla bicicletta nella costruzione di macchine efficienti e leggere per la produzione e la trasmissione di energia. Il premio Kramer non è stato ancora vinto, ma rappresenta una meta che molto probabilmente verrà raggiunta.

Anche la tecnologia della produzione deve molto all'improvvisa richiesta creata dalla bicicletta per un gran numero di pezzi di precisione. La bicicletta di tipo medio è costituita complessivamente da circa 1000 pezzi singoli. Si sa che circa la metà di questi pezzi si trovano nella catena, ma i rimanenti richiedono livelli elevati nella tecnica di stampaggio e di lavorazione meccanica e i metodi messi a punto per produrli hanno rappresentato un grosso passo avanti.

Tra gli accessori che fecero la loro comparsa negli anni 80 e 90 del secolo scorso vi è il dispositivo della ruota libera, adottato inizialmente per facilitare la manovra di salire sulla bicicletta. L'adozione della ruota libera creò la necessità di freni migliori. Originariamente i freni della bicicletta erano del semplice tipo « a cucchiaio » agenti sul pneumatico anteriore. I freni che agiscono sui cerchioni vennero dopo, prima il freno a staffa che agisce sulla superficie interna del cerchione, poi a ganasce, che preme sui lati esterni piatti del sottile cerchione usato nelle biciclette da corsa e nelle biciclette leggere da turismo. Lo scopo iniziale del freno a ganasce era quello di facilitare la sostituzione della ruota senza rimuovere i freni. Questo tipo di freno è fatto funzionare da un cavo in cui un filo flessibile interno in tensione è contenuto in una guaina esterna in compressione. L'efficacia di questo sistema va oltre lo scopo per il quale è stato originariamente messo a punto. Da allora infatti è stato largamente usato per funzioni diverse come, per esempio, nel comando della frizione e dell'acceleratore nelle motociclette, e nei dispositivi di controllo degli aerei.

Successivamente fecero la loro comparsa due altri tipi di freni. Uno è il freno a contropedale, particolarmente comune negli Stati Uniti. L'altro è il freno a tamburo usato oggi nelle automobili e nelle motociclette. Entrambi questi tipi di freni hanno il vantaggio di mantenere la loro efficacia anche in condizioni di pioggia.

Il desiderio di essere in grado di variare il rapporto di trasmissione di una bicicletta per pedalare in salita o contro vento è molto ovvio. Questo rapporto di trasmissione è ancora definito dal diametro di una ruota equivalente a quella del primo velocipede. Di conseguenza una ruota di 68 centimetri di diametro spinta da ruote dentate di 44 e 18 denti ha un « rapporto » di (68 × 44)/18, ossia 166 centimetri. L'idea di avere ingranaggi di differenti dimensioni sulla ruota posteriore portò al cambio di velocità « dérailleur » del 1899, con il quale il ciclista può trasferire la catena da una ruota dentata a un'altra mentre pedala. La variazione della lunghezza della catena è controllata da una puleggia ausiliaria, che funge da tendicatena. Questo tipo di cambio è leggero ed efficace ma necessita di adeguati aggiustamenti e di lubrificazione se si vuole che duri.

Un'alternativa è costituita dal cambio nel mozzo Sturmey-Archer (*si veda l'illustrazione a pagina 37*). In questo sistema è usato un singolo treno di ingranaggi epicicloidali in modo tale che quando la bicicletta si trova nella marcia piú alta, la gabbia è spinta dalla ruota dentata e la ruota dalla corona planetaria. Quando la bicicletta si trova nella marcia intermedia, la spinta è diretta alla ruota, e quando si trova nella marcia piú bassa, la ruota dentata spinge la corona planetaria e la gabbia spinge la ruota a una velocità ridotta. Se per esempio, il pignone (che è fissato all'asse stazionario) ha lo stesso numero di denti dei satelliti e la corona planetaria ne ha tre volte questo numero, allora la marcia alta avrà un rapporto moltiplicante di 4/3 confrontato con la trasmissione diretta e la marcia bassa un rapporto demoltiplicante di 3/4. È disponibile anche una versione a cinque velocità che incorpora due treni epicicloidali a differenti rapporti. Il vantaggio del cambio nel mozzo è la compattezza e il fatto che il meccanismo è ben protetto contro la polvere e i danni.

Sorsero molte ditte per la produzione di parti e accessori per biciclette. Una ditta in particolare, la Joseph Lucas Limited, deve la sua importanza iniziale alla produzione del primo fanale per bicicletta efficiente, una lampada a olio con stoppino nota con il nome di « King of the Road Cycle Hub Lamp ». Successivamente questa compagnia costruí fanali ad acetilene e poi fanali elettrici, cosí che si trovò in posizione favorevole a incrementare il mercato dei fanali per automobili e altri dispositivi elettrici per automobile. Un primo tipo di generatore elettrico per biciclette fu inventato da Richard Weber di Lipsia nel 1886. L'attuale generatore applicato alla forcella, che richiede un minimo sforzo supplementare da parte del ciclista, fu prodotto per la prima volta nel 1936 da Raleigh.

La fabbricazione di biciclette è ancora un grosso affare, con una produzione mondiale tra i 35 e i 40 milioni di veicoli all'anno. Il paese con la produzione maggiore è ancora gli Stati Uniti con circa 6 milioni di biciclette all'anno, seguito da vicino dalla Cina con circa 5 milioni. In ogni caso queste cifre dimostrano l'importanza della bicicletta. Se si esamina la quantità e la

La catena da bicicletta è passata attraverso diversi stadi di sviluppo prima di arrivare alla forma definitiva, la catena a rulli e boccole inventata da Hans Renold nel 1880. La progenitrice era la catena a pioli o a perni (*a*), in cui i pioli gravavano direttamente sui denti delle ruote dentate e le piastrine di collegamento giravano sui perni a ciascuna estremità dei pioli. In una catena di questo tipo si ha un logoramento e un attrito esagerati sia nei denti sia nei fori delle piastrine. Un miglioramento successivo fu la catena a rulli (*b*), in cui l'attrito e il logoramento delle ruote dentate e ridotto da rulli cavi montati sui pioli, ma il logoramento delle piastrine sui perni era ancora troppo alto. L'invenzione di Renold della catena a rulli e boccole (*c*) che distribuiscono il carico sull'intera lunghezza del piolo, superò questo ultimo inconveniente.

L'uomo in bicicletta, in termini di energia consumata nel percorrere una certa distanza in funzione del proprio peso, sale al primo posto nella graduatoria relativa ad animali e macchine in grado di muoversi. La quantità di energia consumata da un ciclista (circa 0,15 calorie per grammo al chilometro) è approssimativamente un quinto di quella consumata da un uomo che cammina (circa 0,75 calorie per grammo al chilometro). Con l'eccezione del punto nero che rappresenta il ciclista (*in basso a destra*) questo grafico è basato su dati compilati da Vauce A. Tucker della Duke University.

Una recente applicazione della tecnologia della bicicletta al volo è rappresentata dall'aereo a spinta umana, una meta che recentemente ha suscitato grande interesse in Inghilterra in seguito all'offerta di un premio di 10 000 sterline da parte di Henry Kramer per il primo volo a spinta umana su un percorso a forma di otto attorno a due pali distanti fra loro circa 800 metri. Finora il volo migliore è stato quello compiuto dal tenente pilota John Potter della Royal Air Force su un monoplano progettato da Christopher Roper. L'aereo di Roper, che si vede nella fotografia con Potter ai comandi, ha un'apertura alare di 25 metri e un peso (senza Potter) di soli 66 chilogrammi. Durante il volo una carenatura racchiude completamente il pilota e i comandi.

distribuzione delle biciclette in uso oggi, si trova che nella maggior parte del mondo hanno una funzione di gran lunga più significativa dell'automobile. La Cina con i suoi 800 milioni di abitanti si affida in maniera massiccia alla bicicletta per il trasporto di persone e di merci. Cosí fanno i paesi del sud-est asiatico e dell'Africa. Perfino l'URSS, con solo circa 1 milione e mezzo di automobili, ha una produzione annuale di biciclette di 4 milioni e mezzo. L'Europa e il Nord America sono quindi in minoranza nel basarsi in maniera cosí eccessiva sulle automobili. Il costo reale di questo comportamento sta diventando sempre piú evidente, non solo nel consumo delle risorse ma anche nell'inquinamento e in altri effetti indesiderabili sulla vita urbana.

Per noi che viviamo nel mondo sovrasviluppato la bicicletta offre una vera alternativa all'automobile, se siamo disposti a riconoscerne e ad afferrarne i vantaggi con l'organizzare il nostro ambiente di vita e di lavoro in modo tale da poter utilizzare questi veicoli. Le eventuali spinte in questo senso sono molte: corsie stradali per biciclette in modo da ridurre per i ciclisti il pericolo del traffico automobilistico, parcheggi per biciclette, facilitazioni per il trasporto delle biciclette su treni e su autobus, e biciclette pubbliche da usare dopo aver parcheggiato l'automobile. Andare in bicicletta è già spesso il modo migliore per spostarsi velocemente nei centri cittadini.

Due fattori importanti devono gradualmente portare a una nuova valutazione della funzione ipertrofica che l'automobile ha nella vita occidentale. Il primo è una indubbia diminuzione delle risorse di combustibile e il conseguente aumento del suo prezzo. Il secondo è l'assoluta ingiustizia nel consumo di energia pro capite tra i paesi che usano le automobili e quelli che non la usano. In questi tempi di comunicazioni universali una tale situazione apparirà sempre piú ingiusta e fonte di disaccordo. È inconcepibile che 800 milioni di cinesi giungano a consumare la stessa quantità di energia pro capite dei 200 milioni di americani; ci deve quindi essere una graduale riduzione nel consumo di energia negli Stati Uniti. A questo scopo la bicicletta può dare un contributo significativo e quindi diventare un grande fattore di livellamento.

Ai paesi in via di sviluppo la bicicletta offre una diversa serie di opportunità. Alla continua diffusione delle biciclette dalle grandi città, a quelle piccole e ai villaggi, si accompagna la diffusione delle capacità meccaniche e delle parti di ricambio essenziali. La tecnologia della bicicletta serve dunque allo sviluppo dell'istruzione tecnica sulla cui base i popoli di questi paesi possono progredire nello stesso modo in cui hanno fatto i paesi sviluppati solo 70-100 anni fa. Vi sono prove di questo processo in atto. I cinesi stanno sostituendo le ruote di legno dei loro tradizionali carretti con ruote da bicicletta rendendoli quindi sia piú facili da spingere sia meno dannosi per le superfici stradali. Le macchine trebbiatrici e vagliatrici sono state progettate in modo da incorporare cuscinetti e trasmissioni a catena del tipo usato per la bicicletta. È questo metodo di « fatelo da soli », è questa tecnologia a livello di villaggio a costituire la strada migliore per un miglioramento autonomo, una strada molto piú plausibile di ogni altra forma di aiuto su larga scala che venga dall'esterno. La potenza della bicicletta rispetto alla potenza della tecnologia moderna piú sofisticata è forse piú evidente in Indocina, dove i nord-vietnamiti la usano come principale mezzo di trasporto. Anche i giapponesi che presero Singapore nella seconda guerra mondiale viaggiarono molto in bicicletta. Ciononostante la bicicletta rimane essenzialmente un mezzo pacifico, e non abbiamo bisogno di includerla nei negoziati per la limitazione delle armi strategiche.

Induzione e probabilità

L'evoluzione del concetto di probabilità e la "crisi dei fondamenti" hanno strettamente legato la teoria della probabilità alla moderna logica induttiva

di Domenico Costantini e Marco Mondadori

Sin dall'antichità è considerato compito della logica analizzare le nostre argomentazioni, cioè analizzare i ragionamenti mediante i quali a partire da certe premesse arriviamo a certe conclusioni. Nel Medioevo si individuarono due tipi di logica: la logica formale, o minore, e quella materiale, o maggiore. La prima avrebbe dovuto occuparsi di analizzare la correttezza delle argomentazioni prescindendo dai loro contenuti. La seconda invece avrebbe dovuto consentire di stabilire le condizioni alle quali i ragionamenti non solo sono corretti ma sono anche veri, arrivano cioè alla scienza « vera ». Dalla fine del secolo scorso però la pretesa di costruire una logica materiale è stata definitivamente abbandonata, almeno dalla maggior parte degli studiosi di logica, di guisa che ora per logica si intende solo la logica formale.

Le nostre argomentazioni possono essere suddivise in due grandi tipi: da una parte quelle che prendono le mosse da premesse vere o presunte tali e consentono di arrivare a conclusioni vere (inferenze dimostrative); dall'altra quelle che non possiedono questa caratteristica, quelle cioè che non conservano la verità delle premesse nel senso che pur partendo da premesse vere o presunte tali non escludono la possibilità di pervenire a conclusioni false (inferenze non dimostrative). La disciplina che studia a livello formale le inferenze dimostrative è nota col nome di logica simbolica o matematica; siamo però convinti che sarebbe piú opportuno chiamarla logica deduttiva. Essa risale sostanzialmente ad Aristotele, fu largamente studiata nel Medioevo e fu sempre nettamente distinta dalla logica materiale.

La disciplina che studia le inferenze non dimostrative è nota come logica induttiva. A differenza di quanto avviene per la logica deduttiva, la distinzione fra logica formale e materiale a livello delle inferenze non dimostrative non è stata affatto netta nel senso che la convinzione che anche la logica induttiva dovesse essere sviluppata come una disciplina formale si è affermata solo molto recentemente se si escludono le idee di G. W. Leibniz a questo riguardo, idee che però restarono del tutto isolate. Infatti il termine « logica induttiva » è stato tradizionalmente usato per designare un insieme di ricerche e di dottrine che hanno la loro origine in F. Bacone e raggiungono il loro massimo sviluppo con J. S. Mill. Nell'ambito di questa tradizione l'induzione viene concepita, appunto da Mill, come un processo inferenziale mediante il quale « concludiamo che ciò che è vero di certi individui di una classe è vero per l'intera classe, o che ciò che è vero in certi tempi sarà vero, in circostanze simili, in tutti i tempi ».

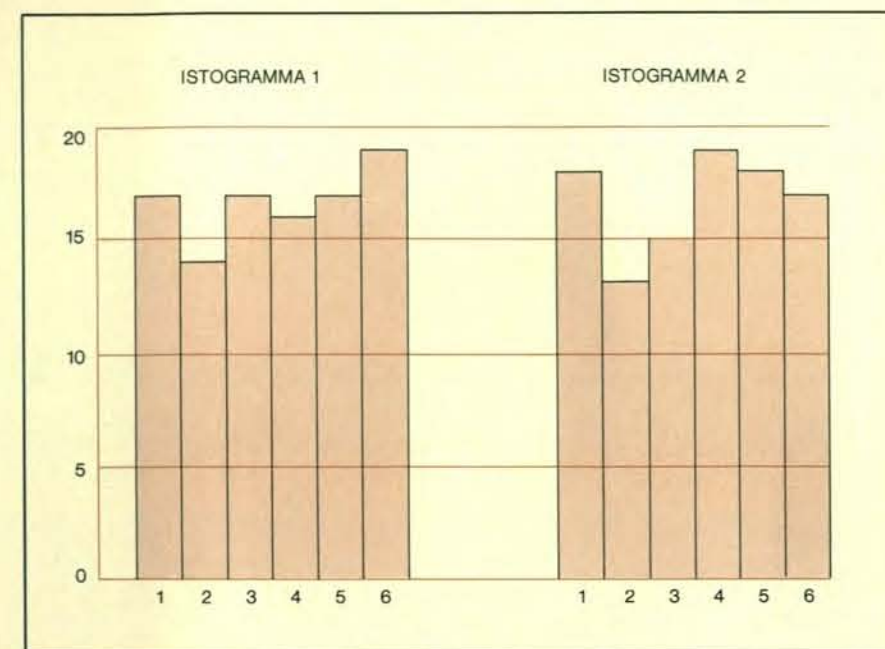
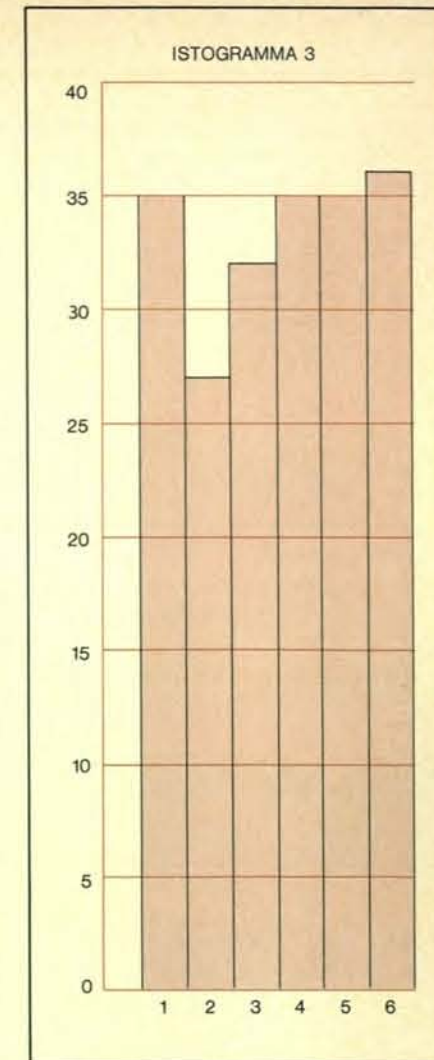
Ma, dato che per l'empirismo classico lo scopo della scienza è la scoperta di leggi generali, ne consegue che la logica induttiva è intesa da Bacone e Mill come la logica della scoperta scientifica, cioè come lo studio e l'esplicitazione dell'insieme di quei metodi che ci portano alla scoperta delle leggi scientifiche, in ultima analisi, come una logica materiale. Ed è appunto in questo quadro che si collocano i famosi quattro metodi milliani della concordanza, della differenza, delle variazioni concomitanti e dei residui.

Induzione e probabilità

Questo modo di intendere il compito e la portata della logica induttiva fu universalmente accettato fino agli inizi del nostro secolo. Come abbiamo detto vi era però un modo alternativo di intendere la logica induttiva anche se esso ha avuto a livello filosofico « ufficiale » un'influenza oltremodo scarsa. Il primo rappresentante di questa ulteriore concezione è – lo abbiamo già visto – Leibniz secondo il quale le leggi generali « non si scoprono mediante l'induzione dagli esempi »; mediante questa induzione possiamo semplicemente giungere ad assegnare un certo grado di probabilità a leggi generali già date. Secondo Leibniz, quindi, la logica induttiva è una disciplina formale nel senso che non è suo compito scoprire le leggi di natura: essa deve limitarsi ad assegnare un grado di probabilità alle leggi di natura delle quali in qualche modo già disponiamo.

Con Leibniz troviamo quindi la prima intuizione del rapporto tra probabilità e induzione, rapporto che costituisce il punto centrale delle moderne ricerche di logica induttiva. Dovette però passare quasi un secolo prima che la geniale intuizione leibniziana riuscisse ad assumere, per opera di P. S. Laplace, la sua prima parziale formulazione scientificamente corretta. Infatti, fu soltanto quando la teoria delle probabilità assunse una veste scientificamente soddisfacente – e ciò avvenne grazie all'opera del grande scienziato francese – che fu possibile disporre di un quadro concettuale adeguato alla formulazione di una concezione della logica induttiva come quella di Leibniz.

Con Laplace viene operata la prima sistemazione rigorosa dei metodi con cui è possibile derivare, da probabilità date di ipotesi, probabilità di ipotesi logicamente connesse con le prime, ipotesi in generale piú complesse di quelle di partenza. Se, per esempio, conosciamo la probabilità dell'ipotesi A e quella dell'ipotesi B, e sappiamo che A e B sono incompatibili, la teoria delle probabilità ci dice che la probabilità dell'ipotesi « A oppure B » è uguale alla somma della probabilità di A e di quella di B. La teoria delle probabilità

FACCE DEL DADO	FREQUENZE RELATIVE PRIMA SERIE	FREQUENZE RELATIVE SECONDA SERIE	FREQUENZE RELATIVE TERZA SERIE
1	0,17	0,18	0,175
2	0,14	0,13	0,135
3	0,17	0,15	0,160
4	0,16	0,19	0,175
5	0,17	0,18	0,175
6	0,19	0,17	0,180

Nei due istogrammi a sinistra sono visualizzati i risultati di due serie di 100 lanci effettuati con un dado comperato in un supermercato qualsiasi. Nel terzo istogramma è invece visualizzato il risultato delle due stesse serie di lanci considerate come un'unica serie di 200 lanci. Le frequenze relative della tabella in alto sono desunte rispettivamente dal primo, secondo e terzo istogramma.

assolve infatti questo compito e, come è bene precisare fin d'ora, *solo* questo compito. Essa non ci dice infatti come debbano essere determinate le probabilità delle ipotesi di partenza; nell'esempio considerato, non ci dice come dobbiamo determinare la probabilità di A e quella di B.

In realtà Laplace si era interessato anche alla determinazione delle probabilità mediante la sua famosa definizione del termine « probabilità » di cui ci occuperemo tra poco. Prima però di analizzare questa definizione è opportuno ribadire un fatto che divenne chiaro solo all'inizio del nostro secolo e al quale abbiamo già fatto cenno quando pocanzi abbiamo detto che la teoria delle probabilità consente solo di passare da probabilità note a probabilità sconosciute legate alle prime. Questo fatto può essere espresso nel modo seguente: la teoria delle probabilità non permette di individuare in modo univoco il significato da attribuire al termine « probabilità » né di formulare alcuna indicazione sul modo con cui le probabilità delle ipotesi devono essere determinate. Questa caratteristica della teoria delle probabilità deve essere sempre tenuta presente perché si tratta di un problema la cui estrema importanza diventa palese quando si pensi che, senza un modo per determinare i valori di probabilità delle ipotesi, non è possibile applicare la teoria delle probabilità.

Contrariamente a quanto potrebbe sembrare a prima vista, il dibattito sul significato da attribuire al termine « probabilità » non è circoscritto a coloro che si occupano professionalmente della teoria delle probabilità. In una certa misura tale dibattito riguarda ciascuno di noi quando, nel servirci di termini quali « probabilità », « probabile » e « probabilmente », ci chiediamo che cosa in realtà intendiamo dire; quasi certamente in casi del genere o non siamo stati in grado di trovare una risposta soddisfacente, oppure ci siamo accorti che la risposta da noi data non era condivisa da altri.

Chi di noi ha avuto l'occasione di cimentarsi con un gioco d'azzardo come la roulette oppure con un concorso a pronostici come il totocalcio – che resta comunque un gioco d'azzardo anche se gestito dallo stato – si sarà certamente chiesto, da una parte, che cosa si intenda dire con l'affermare che il rosso è tanto probabile quanto il nero o che il banco ha maggiori probabilità di vincere rispetto al giocatore; d'altra parte, cosa si intenda quando si sostiene che probabilmente la squadra A vincerà la partita che giocherà la domenica seguente contro la squadra B.

La risposta piú immediata che si può dare nel caso della roulette è che il rosso è tanto probabile quanto il nero perché vi sono 18 numeri rossi e 18

Simbolo	Significato
$\neg$	non
$\vee$	o
$\wedge$	e
$\rightarrow$	se ... allora
$\leftrightarrow$	se e solo se
$\exists$	esiste almeno un
$\forall$	per tutti
$a_1, a_2, \ldots, a_n$	costanti individuali
$P_1, P_2, \ldots, P_m$	costanti predicative

Alcuni dei principali simboli, e relativo significato, appartenenti al linguaggio simbolico di cui si parla nell'articolo. Le costanti individuali denotano individui descritti dal linguaggio mentre le costanti predicative descrivono proprietà godute dagli individui.

numeri neri e quindi i casi favorevoli al rosso sono tanti quanti quelli favorevoli al nero. Parimenti la ragione per cui il banco ha maggior probabilità di vincere di quante non ne abbia il giocatore va cercata nel fatto che quest'ultimo puntando, per esempio, sui numeri pari ha soli 18 casi favorevoli mentre il banco ne ha 19, cioè 18 dispari e lo 0.

La risposta data da uno scommetitore del totocalcio circa il significato da attribuire al termine « probabilmente » sarà sicuramente diversa. Essa infatti potrà essere che la squadra *A* vincerà probabilmente la partita della domenica seguente perché i tre migliori giocatori della squadra *B* sono stati squalificati e quindi non potranno giocare; oppure la risposta potrebbe anche essere che la squadra *B* sta attraversando un periodo di scarsa forma come è dimostrato dal fatto che da cinque domeniche non riesce a vincere una partita, ecc.

È evidente che le due spiegazioni ora viste del termine « probabilità » sono diametralmente opposte. In un caso infatti si fa appello all'analisi dei casi favorevoli all'ipotesi che interessa, mentre nell'altro si ricorre a certe informazioni piú o meno « soggettive » di cui il tifoso dispone.

Le varie accezioni di « probabilità »

La disparità di opinioni che abbiamo incontrato a livello del modo comune di parlare la ritroviamo anche a livello scientifico, come vedremo subito esponendo le varie definizioni di probabilità che sono state proposte. Cominciamo questa esposizione dalla definizione classica di probabilità, definizione dovuta, come abbiamo detto, a Laplace. Il grande scienziato francese, quando per la prima volta introdusse esplicitamente una definizione di probabilità, argomentò in modo del tutto simile a quello del nostro giocatore di roulette. Ciò non deve meravigliare dal momento che la prima formulazione scientificamente accettabile della teoria delle probabilità operata da Laplace nei primi anni del secolo scorso costituiva la razionalizzazione di una serie di risultati, relativi ai fenomeni casuali, che erano andati accumulandosi nel XVIII secolo e che per la maggior parte erano relativi ai giochi d'azzardo. Come abbiamo visto nel nostro esempio della roulette, in questi casi è del tutto naturale legare la probabilità di un evento ai casi favorevoli all'evento stesso (parlare di probabilità di un'ipotesi o, come abbiamo fatto ora, di probabilità di un evento, non comporta alcuna differenza sostanziale: si può infatti immediatamente passare dalle une alle altre prendendo in considerazione rispettivamente l'ipotesi che descrive l'evento e l'evento descritto dall'ipotesi). Gli scienziati che per primi nel XVIII secolo si interessarono a questi problemi, e Laplace che diede una sistemazione rigorosa ai loro risultati, definiscono la probabilità di un evento come il rapporto fra il numero dei casi favorevoli al verificarsi dell'evento stesso e il numero dei casi possibili purché questi ultimi siano tutti ugualmente possibili.

Questa definizione di probabilità fu unanimemente accettata e, prescindendo da qualche isolata voce discorde, fu la definizione adottata durante tutto il secolo scorso. Abbiamo però detto che qualche studioso, soprattutto nella seconda metà del XIX secolo, non era affatto d'accordo con l'opinione piú diffusa. Questi dissensi diventarono consistenti nei primi anni del nostro secolo e gli argomenti contrari alla definizione laplaciana possono essere riassunti nel modo seguente.

L'applicazione di questa definizione dipende in modo essenziale dalla clausola secondo cui i casi possibili debbono essere tutti ugualmente possibili. Ma come facciamo a sapere che un dato insieme di casi possibili è un insieme di casi equipossibili? La risposta di Laplace è la seguente: « La teoria della probabilità consiste nel ridurre tutti gli eventi di un certo tipo a un certo numero di casi ugualmente possibili, *cioè tali che possiamo essere ugualmente incerti circa la loro esistenza* ». La frase in corsivo è una delle possibili formulazioni di un famoso principio enunciato da J. Bernoulli sotto il nome di principio di ragione non sufficiente e ora noto invece come principio d'indifferenza. Per la maggior parte delle situazioni induttive del tipo considerato per lo piú da Laplace, risulta non solo abbastanza facile, ma anche abbastanza naturale analizzare la situazione in un insieme di casi equipossibili, tali cioè che non sussistano ragioni di credere che se ne verificherà uno di essi piuttosto che un altro (il lettore si sarà accorto che questa è un'altra formulazione del principio d'indifferenza). In effetti nei casi dei giochi d'azzardo che fanno uso di meccanismi non truccati (dadi, monete, roulette, urne) è molto plausibile l'ipotesi di equipossibilità dei casi possibili. Ma le difficoltà si presentano non appena abbiamo qualche ragione di sospettare che i dadi, le monete, le roulettes sono truccati. Per esempio nel caso di un dado zavorrato è ovvio che le uscite di ciascuna delle sei facce non potranno piú essere considerate come equipossibili. A un'analisi piú sottile la definizione di Laplace si rivela già inapplicabile, almeno in senso generale, agli stessi casi per i quali era stata formulata.

Al di fuori di questi casi l'inapplicabilità di tale definizione è addirittura palese: che cosa significa infatti che il signor Rossi che ha compiuto 38 anni ha una probabilità pari a 0,003 di morire nel corso del 39° anno di vita? Significa forse che il nostro signore ha di fronte 1000 modi alternativi di vivere di cui solo 3 lo conducono a morte nel corso del prossimo anno? L'assurdità di una simile ipotesi è cosí evidente da non meritare alcun commento.

Le critiche che abbiamo ricordato sono dovute a R. von Mises che le formulò attorno al 1920 e segnarono l'inizio del periodo ormai noto come « crisi dei fondamenti della probabilità », cioè di quel periodo di dibattiti volti a

individuare una nuova spiegazione del significato del termine « probabilità » da sostituire a quella data da Laplace all'inizio del XIX secolo.

Von Mises fece seguire alle critiche viste una nuova definizione di probabilità; serviamoci di un esempio per meglio afferrare la innovazione contenuta nella definizione da lui proposta. Supponiamo di lanciare un certo numero di volte un dado e di registrare i risultati dei nostri lanci nel seguente prospetto:

1°	2°	3°	4°	5°	6°	7°
6	4	2	1	6	4	2
8°	9°	10°	11°	12°	13°	...
1	3	2	3	1	1	...

Chiamiamo ora *frequenza relativa* dell'evento « uscita di 1 » il numero di volte in cui compare l'1 in un certo numero di lanci diviso il numero di questi ultimi. Nel nostro esempio la frequenza relativa di questo evento sarà allora 4/13 mentre la frequenza relativa dell'evento « uscita di 6 » sarà 2/13. Ciascun evento avrà una sua frequenza relativa che varierà col variare del numero di lanci che prendiamo in considerazione. La frequenza relativa dell'evento « uscita di 1 », quando ci limitiamo a considerare i primi sei lanci, sarà 1/6.

Von Mises sostenne che per probabilità del verificarsi di un certo evento in una successione casuale di eventi simili e ripetibili, quali i lanci di un dado del nostro esempio, doveva intendersi il limite al quale tende la frequenza relativa di quell'evento quando il numero dei termini della successione cresce indefinitamente, ovviamente quando questo limite esiste. È bene precisare che il termine « limite » usato in questo contesto non è il preciso concetto dell'analisi matematica bensí il concetto impreciso che ciascuno di noi usa quando nel linguaggio di tutti i giorni si serve di questo termine.

All'incirca nello stesso periodo fu però proposta, a opera dell'economista J. M. Keynes, una terza definizione di probabilità. Per Keynes la probabilità di una ipotesi è una relazione logica che lega un insieme di informazioni disponibili per un certo individuo o per una certa collettività e l'ipotesi stessa. Questo autore non precisava tuttavia che cosa dovesse intendersi per questa relazione logica cosí che l'intera definizione restava alquanto oscura. Si può dare un'idea di ciò che Keynes intendeva facendo intervenire una relazione logica per spiegare il significato della probabilità, tornando all'esempio del tifoso che riteneva probabile la vittoria della squadra *A* nella partita della domenica seguente. Il nostro tifoso riteneva infatti che esistesse qualche relazione oggettiva, e in questo senso logica, tra la mancanza dei tre migliori giocatori di *B* e il rendimento di questa squadra nella prossima partita, oppure che lo stato di scarsa forma dimostrato dalla squadra *B* nelle ultime cinque partite fosse in qualche modo oggettivamente connesso con il comportamento di questa squadra nella prossima partita con la squadra *A*. In breve, il nostro tifoso riteneva che la vittoria della squadra *A* fosse probabile perché un insieme di informazioni in suo possesso sostenevano l'ipotesi della vittoria di *A* piú di quanto sostenessero l'ipotesi contraria.

Alcuni anni piú tardi, attorno al 1930, due autori, l'inglese F. P. Ramsey e l'italiano B. de Finetti, pur partendo da considerazioni del tutto differenti, arrivarono a una medesima conclusione relativamente al significato da attribuire al termine « probabilità », conclusione del tutto diversa da quella alla quale erano arrivati i due autori che abbiamo già preso in esame. Ramsey infatti, dopo aver tentato senza successo di dare un significato alla relazione logica che compare nella definizione di Keynes, si convinse che l'unica possibilità di legare fra di loro le informazioni in possesso di un individuo e una ipotesi da questi avanzata, consisteva nel fare appello alla fiducia che l'individuo nutriva nella verità della ipotesi in questione. De Finetti invece, muovendo dalla convinzione che ogni evento è un fatto unico non ripetibile e che, conseguentemente, non ha senso ipotizzare la ripetibilità degli eventi come presuppone la definizione di von Mises, arrivò come Ramsey a sostenere che la probabilità di un evento è la fiducia che un dato individuo nutre nel verificarsi dell'evento stesso.

Una conseguenza immediata di questi contrastanti punti di vista era che ciascuno di essi proponeva un modo,

1. Assioma dell'equivalenza
Se $\vdash H \longleftrightarrow H'$ e $E \longleftrightarrow E'$, allora $c(H, E) = c(H', E')$.
2. Assiomi dell'invarianza
PER LE COSTANTI INDIVIDUALI:
Se $q(H)$ e $q(E)$ sono ottenuti mediante q da H ed E, allora $c(H, E) = c(q(H), q(E))$.
PER LE COSTANTI PREDICATIVE:
Se $r(H)$ e $r(E)$ sono ottenuti mediante r da H ed E, allora $c(H, E) = c(r(H), r(E))$.
PER LE FAMIGLIE:
Se $t(H)$ e $t(E)$ sono ottenuti mediante t da H ed E, allora $c(H, E) = c(t(H), t(E))$.
3. Assioma della rilevanza degli esempi positivi
Se a_i e a_j sono costanti individuali che non compaiono in E, allora $c(P_n a_i, E \wedge P_n a_j) > c(P_n a_i, E)$.

Rappresentazione simbolica degli assiomi della logica induttiva di cui si parla nell'articolo. c (H, E) significa la probabilità di H sulla base di E; nei casi interessanti l'enunciato H descrive un'ipotesi relativa a un fatto sconosciuto mentre l'enunciato E descrive un insieme di risultati sperimentali conosciuti da chi formula l'ipotesi. Il simbolo $\vdash$ serve per indicare che quanto lo segue è dimostrabile, e quindi $\vdash H \longleftrightarrow H'$ significa che H e H′ sono logicamente equivalenti. q, r e t sono permutazioni rispettivamente delle costanti individuali, delle costanti predicative e delle famiglie con lo stesso numero di membri del linguaggio simbolico.

diverso da quello proposto dagli altri, di determinare le probabilità. Per i frequentisti, cioè per coloro che accettano la definizione di von Mises, la probabilità deve essere determinata sulla base di osservazioni sperimentali, cioè esaminando le frequenze relative con cui gli eventi si verificano. Per i logicisti, cioè per coloro che accettano la tesi di Keynes, la probabilità deve essere determinata cercando di dare un valore numerico alla relazione logica che lega le conoscenze di cui si dispone e l'ipotesi che descrive il verificarsi dell'evento. Per i soggettivisti infine, cioè per coloro che fanno proprie le argomentazioni di Ramsey e de Finetti, le probabilità devono essere determinate rendendo esplicita mediante l'uso di scommesse la fiducia degli individui nel verificarsi dell'evento. È chiaro che, mentre per i frequentisti e i logistici vi è una sola probabilità per ogni evento, nel senso che la probabilità è un fatto esterno all'uomo che questo deve scoprire, per i soggettivisti invece ogni uomo può assegnare a un dato evento una sua propria probabilità e il fatto che due individui possano avere la stessa fiducia nel verificarsi di un evento e quindi assegnargli la stessa probabilità, è un fatto del tutto accidentale.

L'assiomatizzazione di Kolmogorov

All'inizio degli anni trenta le cose stavano quindi come abbiamo visto: esistevano cioè tre modi contrapposti di definire la probabilità, e di conseguenza tre modi diversi di determinare la probabilità; ma ancora, ciò che è senza dubbio più grave, si credeva che esistessero tre teorie diverse delle probabilità poiché ciascuno degli indirizzi che abbiamo passato in rassegna sosteneva che la teoria delle probabilità era quella sviluppata a partire dalla definizione che esso sosteneva, cioè era rispettivamente la teoria delle successioni casuali di eventi, delle relazioni logiche tra evidenze e ipotesi e della fiducia degli individui nel verificarsi degli eventi.

La crisi dei fondamenti della probabilità si rifletteva quindi sulla stessa teoria delle probabilità con evidente grave pregiudizio di quest'ultima. Questo stato di incertezza fu però superato nel 1933 quando il matematico sovietico A. N. Kolmogorov dimostrò che la teoria delle probabilità poteva essere completamente sviluppata senza mai farvi intervenire il significato del termine « probabilità ». In altre parole, il matematico sovietico mostrò che la teoria delle probabilità – come del resto tutta la matematica di cui questa teoria è una branca – si occupa di entità astratte che non necessitano, almeno a livello dello sviluppo della teoria, di alcuna interpretazione. Kolmogorov si poneva cioè sulle orme di D. Hilbert, e, come quest'ultimo aveva mostrato che l'intera geometria poteva essere sviluppata senza fare riferimento al significato dei termini « punto », « retta », « piano », ecc., così Kolmogorov mostrò che l'intera teoria delle probabilità può essere sviluppata senza fare riferimento al significato del termine « probabilità », limitandosi a introdurre assiomaticamente alcune sue caratteristiche, per altro molto generali, note appunto come assiomi della teoria della probabilità: *a)* la probabilità è una funzione i cui argomenti sono eventi (ipotesi) e i cui valori sono numeri reali maggiori o uguali a zero e minori o uguali a uno; *b)* la probabilità di due eventi (ipotesi) che non possono verificarsi contemporaneamente (che sono incompatibili) è la somma delle probabilità dei due eventi (ipotesi).

L'opera di Kolmogorov ebbe una importanza notevolissima innanzitutto perché, liberando la teoria delle probabilità dalle dispute relative ai fondamenti, le permise uno sviluppo indipendente e rigoglioso tanto che oggi questa disciplina rappresenta una delle branche della matematica più attive e feconde di risultati. Tuttavia l'influenza del fondamentale lavoro del matematico sovietico non restò limitata a ciò: la generalizzazione dell'istanza di fondo che caratterizza il lavoro di Kolmogorov avrebbe infatti portato a una svolta radicale nella crisi dei fondamenti della probabilità, cioè alla nascita della logica induttiva moderna.

Abbiamo detto che la tesi di fondo su cui si basa il lavoro del matematico sovietico è la seguente: la teoria delle probabilità può e deve essere sviluppata senza fare riferimento al significato del termine « probabilità ». Vediamo ora cosa significhi generalizzare questa tesi. È chiaro che, prescindendo dalle intenzioni sostanzialmente metafisiche volte a individuare l'essenza della probabilità, intenzioni che ora non ci interessa analizzare, il significato delle ricerche connesse con la crisi dei fondamenti della probabilità risiede nella convinzione che solo quando si è attribuito un significato al termine « probabilità » si possano determinare i valori di probabilità e quindi in ultima istanza si possa applicare la teoria delle probabilità. È ora naturale, accettando la istanza formalistica di Kolmogorov, porsi la seguente domanda: se è possibile sviluppare la teoria delle probabilità senza fare appello al significato del termine « probabilità », non sarà possibile trovare il modo di determinare anche i valori di probabilità senza fare riferimento al significato di questo termine? Se si generalizza l'impostazione formalistica di Kolmogorov, non si può non dare risposta affermativa a questa domanda; ciò significa tentare di elaborare dei metodi atti alla determinazione delle probabilità nei quali non si faccia alcun riferimento a un qualunque significato di « probabilità ». La domanda successiva sarà allora come ciò sia possibile. La nostra risposta è che le probabilità vengono di fatto determinate – in generale ma non necessariamente – mediante i cosiddetti metodi statistici. Esempi di questa determinazione si hanno nelle ricerche fisiche, astronomiche, economiche, demografiche, genetiche, farmacologiche, meteorologiche e in molte altre discipline, in sostanza in tutte le discipline sperimentali.

La logica induttiva

Le ricerche relative ai fondamenti della probabilità acquistano quindi un significato nuovo e completamente diverso da quello che hanno avuto nella prima metà del secolo: non si tratterà più di andare alla ricerca di esplicazioni del termine « probabilità » bensì dei principi che di fatto regolano la determinazione delle probabilità nelle scienze sperimentali. Una volta che questi principi siano stati individuati, dovranno essere formalizzati e, sulla base di questa formalizzazione, se ne dovrà studiare la portata, la compatibilità reciproca, l'indipendenza reciproca e così di seguito; se ne dovrà cioè analizzare la struttura logica e le implicazioni epistemologiche.

Alcuni di questi principi sono già stati individuati e dopo essere stati formalizzati sono stati assunti quali assiomi della logica induttiva; ora ne esamineremo alcuni fra i più importanti. La logica induttiva costituisce un ampliamento della teoria delle probabilità; come abbiamo detto, quest'ultima può a pieno diritto essere considerata una parte della logica induttiva. Gli assiomi di questa disciplina comprenderanno quindi anche quelli della teoria delle probabilità; ciò significa che una volta determinati i valori di probabilità delle ipotesi induttive, essi possono essere « maneggiati » con i teoremi della teoria delle probabilità.

Abbiamo visto che Kolmogorov basò il suo sistema d'assiomi sul presupposto che la probabilità fosse una funzione: ciò ci assicura l'esistenza di uno e un solo valore di probabilità in corrispondenza di ogni ipotesi. Sia quindi p la funzione-probabilità; essa avrà come argomenti enunciati (ipotesi) di un certo linguaggio simbolico e come valori numeri reali non negativi. Se H è un enunciato, indicheremo con $p(H)$ (a volte anche con p_H) il valore che la funzione-probabilità assume in corrispondenza di H (la probabilità di H).

Siano H e H' enunciati, imponiamo allora che la probabilità soddisfi le seguenti condizioni:

assioma dell'equivalenza: se H è logicamente equivalente a H', allora $p(H) = p(H')$;

assioma del limite superiore: se H è logicamente certo, allora $p(H) = 1$;

assioma dell'additività: se H e H' sono logicamente incompatibili, allora $p(H \vee H') = p(H) + p(H')$.

Ci limiteremo a illustrare l'assioma dell'equivalenza poiché, come il lettore avrà certamente notato, degli altri due ci siamo già occupati esaminando l'assiomatizzazione di Kolmogorov. Il lettore si sarà inoltre accorto che tra gli assiomi allora visti non compariva quello dell'equivalenza; ciò dipende dal fatto che Kolmogorov costruisce il suo sistema assiomatico sulla nozione di « evento » e non su quella di « enunciazione » come noi stiamo facendo, e questo ci impone di prendere delle precauzioni in vista di evitare spiacevoli conseguenze. L'assioma dell'equivalenza afferma infatti che ipotesi logicamente equivalenti non possono avere probabilità diverse; in questo modo evitiamo l'eventualità senza dubbio spiacevole di assegnare probabilità diverse a ipotesi con lo stesso contenuto informativo, a ipotesi cioè che « dicono la stessa cosa » seppure in modi diversi.

Dopo aver esaminato gli assiomi della teoria delle probabilità passiamo a un assioma che sta per così dire a mezza strada fra la teoria delle probabilità e la logica induttiva. Si tratta cioè di un assioma la cui validità alcuni ritengono debba essere confinata alla logica induttiva mentre altri vorrebbero estendere anche alla teoria delle probabilità. Esso fu introdotto dai logici induttivi ma le ricerche di cui è stato oggetto hanno consentito di metterne in luce aspetti indubbiamente probabilistici. Ma passiamo all'enunciazione dell'*assioma di regolarità*: se $p(H) = 1$, allora H è logicamente certo. Questo assioma è indubbiamente legato a quello del limite superiore: questo ci assicura che la probabilità delle ipotesi certe è pari a uno, quello di regolarità che vale anche l'inverso. La loro congiunzione porta a questa conclusione: le ipotesi certe sono tutte e sole quelle la cui probabilità è pari a uno.

Ma passiamo alla logica induttiva in senso proprio concentrando la nostra attenzione su alcuni fra i più caratteristici assiomi di questa disciplina. Intendiamo riferirci agli assiomi dell'invarianza, che per certi aspetti si ricollegano al principio d'indifferenza di Bernoulli, e all'assioma di rilevanza degli esempi positivi.

Gli assiomi dell'invarianza sono largamente accettati dalla metodologia statistica e possono trovare una matrice comune nel principio secondo cui, in alcune circostanze, quello che ha importanza nella valutazione delle probabilità è la quantità delle osservazioni e non anche la loro qualità. Ma cerchiamo di chiarire questa affermazione con alcuni semplici esempi. Come è d'uso in questi casi, supponiamo di voler determinare la probabilità di estrarre una pallina di un dato colore da un'urna che sappiamo contenerne di bianche e di nere in proporzione sconosciuta. Supponiamo inoltre di operare la determinazione dopo aver fatto un certo numero di estrazioni, cioè di valutare la probabilità di un'estrazione futura sulla scorta della conoscenza dei risultati di alcune estrazioni passate. Ovviamente l'esempio di cui ci serviremo può essere facilmente generalizzato sia al caso in cui i colori delle palline contenute nell'urna siano più di due sia al caso in cui si tratti di esaminare individui provenienti da una popolazione qualsiasi.

Supponiamo quindi di aver estratto n palline dall'urna e di aver costatato che n_B sono bianche e le restanti n_N sono nere. Il nostro problema sarà allora di determinare la probabilità di estrarre al prossimo colpo una pallina bianca o, ciò che sostanzialmente è lo stesso, di determinare la probabilità dell'ipotesi secondo cui la prossima pallina che estrarremo sarà bianca. Supponiamo di averlo risolto in qualche modo cioè usando un metodo che però ora non ci interessa analizzare. In prima istanza l'assioma dell'invarianza che per primo prenderemo in esame afferma che questa probabilità dipende unicamente da n_B e n_N e non anche dall'ordine con cui le palline bianche e nere si sono presentate nel corso delle n estrazioni già eseguite. Se cioè $n_B = 2$ e $n_N = 1$, le tre estrazioni che possono aver dato luogo a questo risultato sono ovviamente le seguenti:

$$a)\ Ba_1Ba_2Na_3$$
$$b)\ Ba_1Na_2Ba_3$$
$$c)\ Na_1Ba_2Ba_3$$

in cui B e N stanno per bianca e nera e a_1, a_2 e a_3 indicano rispettivamente il risultato della (la pallina estratta alla) prima, seconda e terza estrazione.

Indichiamo con p_{Ba}, p_{Bb} e p_{Bc} le probabilità di estrarre al prossimo colpo una pallina bianca calcolata supponendo noti rispettivamente i risultati *a)*, *b)* e *c)*. L'assioma in questione stabilisce che

$$p_{Ba} = p_{Bb} = p_{Bc};$$

questa probabilità la indicheremo con p_B.

Se ora formalizziamo il nostro discorso traducendo *a)*, *b)* e *c)* in un opportuno linguaggio simbolico, B e N diventeranno segni per proprietà (costanti predicative) e a_1, a_2 e a_3 segni per individui (costanti individuali). In questo caso l'assioma di cui ci stiamo occupando diventa l'assioma *dell'invarianza relativamente alle costanti individuali*: la probabilità di un enunciato è invariante relativamente alla permutazione delle costanti individuali del linguaggio. Infatti i tre risultati sperimentali visti e anche gli enunciati che nel linguaggio simbolico li descrivono, si ottengono l'uno dall'altro permutando opportunamente gli individui (le costanti individuali). Una diversa formulazione di questo assioma, che può essere più convincente a livello intuitivo, è la seguente: nel valutare le probabilità tutti gli individui (le costanti individuali) devono essere trattati alla pari.

Prima di passare al secondo e al terzo assioma dell'invarianza, introduciamo il concetto di *famiglia* di proprietà (costanti predicative); per famiglia intenderemo un insieme di proprietà (costanti predicative) a due a due incompatibili e tali che ogni individuo (costante individuale) deve godere di almeno una di esse. Le proprietà coinvolte nelle estrazioni che abbiamo finora preso in considerazione danno origine a una famiglia; infatti una pallina non può essere bianca e nera e ogni pallina che estraiamo gode di una delle due proprietà, questo ovviamente per il modo con cui abbiamo costruito il dispositivo sperimentale. Anche le proprietà « essere testa » ed « essere croce » riferite ai possibili risultati del lancio di una moneta sono una famiglia. Infatti lanciando una moneta non possono presentarsi contemporaneamente due facce e d'altra parte una delle due facce deve presentarsi, ovviamente quando si escluda che la moneta possa fermarsi in bilico sullo spigolo. Ci si convince facilmente che anche le date di nascita degli individui e il luogo in cui si trovano in un certo istante sono famiglie così come le loro stature, i loro perimetri toracici e così di seguito.

Supponiamo ora d'aver fatto una nuova serie di estrazioni tali che $n_B = 1$ e $n_N = 2$. Come sappiamo le tre estra-

zioni che possono aver dato luogo a questo risultato sono le seguenti:

d) $Na_1Na_2Ba_3$

e) $Na_1Ba_2Na_3$

f) $Ba_1Na_2Na_3$

in cui i simboli hanno il significato che conosciamo. Supponiamo ora di essere in grado di determinare le probabilità che la pallina estratta al prossimo colpo sia nera sulla base della conoscenza di *d*), *e*) e *f*) e inoltre che esse siano

$$p_{Nd} = p_{Ne} = p_{Nf}.$$

Quest'ultima uguaglianza vale ovviamente in virtú dell'assioma dell'invarianza relativamente alle costanti individuali; indichiamo questa probabilità con p_N. Il secondo assioma dell'invarianza stabilisce che

$$p_N = p_B,$$

cioè che la probabilità di estrarre una pallina nera, quando si sappia che sono state estratte due palline nere e una bianca, deve essere uguale alla probabilità di estrarre una pallina bianca, quando si sappia che sono state estratte due palline bianche ed una nera. Generalizzato e tradotto a livello simbolico esso diventa l'assioma dell'invarianza relativamente alle costanti predicative: la probabilità di un enunciato è invariante relativamente alla permutazione delle costanti predicative di una famiglia. Anzi, in questo caso vi è un'altra formulazione: nel valutare le probabilità tutte le proprietà (costanti predicative) di una famiglia devono essere trattate alla pari.

Prendiamo ora in considerazione, oltre a quello delle urne, un nuovo esperimento, supponiamo cioè di aver lanciato tre volte una moneta senza conoscere se la stessa è truccata oppure no, e di aver osservato che due volte si è presentata la faccia con la croce e una volta quella con la testa, cioè $n_C = 2$ e $n_T = 1$. Supponiamo inoltre di essere in grado di valutare la probabilità che nel prossimo lancio si presenti la faccia con la croce e che essa sia pari a p_C. Se applichiamo il principio che finora ci ha guidato dovremo concludere che, pur essendo il lancio di una moneta e l'estrazione dall'urna esperimenti che coinvolgono proprietà diverse, debba valere

$$p_C = p_B$$

in virtú del fatto che sia nell'uno sia nell'altro caso i risultati favorevoli all'ipotesi erano due e quello contrario uno. L'assioma che ha diretto le nostre considerazioni è l'assioma *dell'invarianza relativamente alle famiglie*: la probabilità di un enunciato è invariante relativamente alla permutazione delle famiglie del linguaggio con lo stesso numero di membri.

L'ultimo assioma della logica induttiva che intendiamo prendere in esame, traduce a livello formale un principio che non solo è largamente usato in tutte le elaborazioni statistiche dei dati sperimentali, ma è anche una delle tesi fondamentali dell'empirismo: si tratta del principio secondo cui *dobbiamo* imparare dall'esperienza.

Si supponga di aver compiuto n estrazioni da un'urna di cui al solito non conosciamo la composizione. Delle n estrazioni effettuate n_B sono state palline bianche. Supponiamo ancora di disporre di un metodo che ci consenta la determinazione delle probabilità e che grazie a questo si sia calcolata pari a p_B la probabilità di ottenere una pallina bianca nella prossima estrazione. Questa probabilità tiene ovviamente conto delle estrazioni effettuate. Supponiamo ora che la situazione sperimentale sia leggermente mutata nel senso che delle n estrazioni fatte non piú n_B siano palline bianche bensí $n_B + 1$; ciò significa che nella nuova situazione sperimentale avremo estratto una pallina bianca in piú rispetto alla situazione presa precedentemente in esame. Sulla base di questo mutato risultato sperimentale calcoliamo, servendoci del metodo di cui disponiamo, la probabilità di estrarre una pallina bianca e sia $'p_B$ questa probabilità. L'assioma *della rilevanza degli esempi positivi* impone che qualunque sia il metodo usato per determinare p_B e $'p_B$ debba valere

$$p_B < {'p_B};$$

cioè: la probabilità di un'ipotesi deve crescere col crescere del numero degli individui osservati che verificano l'ipotesi stessa.

Con questo assioma abbiamo concluso il nostro esame della logica induttiva; ricordiamo però ancora una volta che gli assiomi di cui ci siamo occupati sono solo una parte di quelli che stanno alla base dei sistemi finora noti di logica induttiva. Ci siamo infatti limitati ad alcuni assiomi che ci sono sembrati particolarmente adatti a corroborare le nostre affermazioni sulla natura e gli scopi della logica induttiva moderna. Siamo inoltre convinti che questo esame sia servito a dare la misura di quanto queste moderne ricerche si riallaccino a quelle che, secondo Leibniz, avrebbero dovuto costituire l'oggetto della logica induttiva.

Indipendentemente comunque dal ricollegarsi o meno alle intuizioni leibniziane, la disciplina di cui abbiamo ora visto alcuni assiomi merita bene il nome di logica induttiva che le abbiamo dato. Infatti, mentre la logica deduttiva si occupa della struttura logica delle inferenze dimostrative, la logica induttiva si occupa con gli stessi intenti delle inferenze non dimostrative, cioè della esplicitazione e dell'analisi delle modalità d'assegnazione delle probabilità alle ipotesi induttive. La logica induttiva moderna può quindi a pieno diritto essere inclusa nella logica formale.

L'opera di Carnap

Abbiamo detto all'inizio del presente articolo che è solo nei primi anni del nostro secolo che la logica induttiva acquista il suo significato moderno e abbandona definitivamente ogni pretesa di porsi come la logica della scoperta scientifica. Questa affermazione, sulla base di quanto abbiamo visto, deve però essere ridimensionata nel senso che è all'inizio del nostro secolo che matura la crisi dei fondamenti della probabilità, crisi che sfocerà nella nascita della moderna logica induttiva. In senso lato, quindi, questa disciplina può farsi risalire all'inizio del nostro secolo solo se si intende – e questa è la nostra opinione – come la naturale conseguenza e il superamento della crisi dei fondamenti della probabilità. In senso stretto, cioè come analisi razionale delle modalità di assegnazione delle probabilità alle ipotesi induttive, essa è il risultato di ricerche portate avanti negli ultimi venti anni principalmente per opera di R. Carnap, il filosofo della scienza tedesco recentemente scomparso.

La logica induttiva quindi è una disciplina molto giovane ed è naturale che stia muovendo i suoi primi passi fra difficoltà d'ogni genere. Ciò nonostante ha già al suo attivo alcuni risultati notevoli, per esempio la dimostrazione della non indipendenza dell'assioma di rilevanza degli esempi positivi e la dimostrazione di un certo tipo di incompatibilità tra l'assioma di invarianza relativamente alle famiglie e gli altri assiomi. In questa sede non possiamo occuparci della non indipendenza dell'assioma della rilevanza a ragione delle difficoltà tecniche che comporta la sua dimostrazione. Il secondo problema può invece essere affrontato, almeno fino a un certo punto, senza fare ricorso ad alcun tecnicismo. Abbiamo parlato di « un certo tipo » di incompatibilità perché non si è dimostrato che da questo assioma, insieme con gli altri della logica induttiva, si può derivare una contraddizione, ma piuttosto che la sua accettazione conduce a un risultato intuitivamente paradossale che può essere evitato se si lascia cadere l'assioma o se ne limita la portata.

Il paradosso della conferma, questo è il nome con cui è noto questo risultato paradossale, può essere formulato nel modo seguente. È opinione comune che se uno scienziato avanzasse la ipotesi « Tutte le pietre di Marte sono sferiche », ipotesi che indicheremo con H, egli dovrebbe, in vista di aumentare il grado di probabilità dell'ipotesi stessa (per confermarla), osservare le pietre di Marte e costatare che sono sferiche. Sarà cioè necessario che il nostro scienziato si ponga in grado di osservare le pietre di quel pianeta allo scopo di vedere se sono effettivamente sferiche. In altri termini, in accordo con l'assioma della rilevanza, per confermare l'ipotesi H sarà necessario osservare oggetti che siano contemporaneamente pietre di Marte e sferici. Ora se ci rammentiamo l'assioma dell'equivalenza giungiamo al paradosso.

L'ipotesi H è logicamente equivalente alla seguente « Tutti gli oggetti non sferici non sono pietre di Marte » che indicheremo con H'. Questa ipotesi è confermata dall'osservazione di oggetti che non sono né sferici né pietre di Marte, per esempio, un autocarro di Pavia o un tram di Milano. Ma H è logicamente equivalente a H' e quindi, per l'assioma dell'equivalenza, la probabilità di H deve essere uguale a quella di H'; ne consegue che crescendo la probabilità di H' deve crescere anche quella di H, in definitiva che l'osservazione di autocarri di Pavia o tram di Milano conferma l'ipotesi che tutte le pietre di Marte sono sferiche. Quindi, per aumentare il grado di probabilità dell'ipotesi relativa alla forma delle pietre di Marte, lo scienziato che l'ha formulata dovrebbe mettersi alla ricerca di oggetti non sferici per vedere se sono o no pietre di Marte. Per confermare la sua ipotesi il nostro scienziato non dovrebbe recarsi su Marte o nelle sue vicinanze perché gli basterebbe controllare che nella sua città esistono un gran numero di oggetti che non sono sferici. Una conferma di questo tipo è indubbiamente inaccettabile, d'altra parte è una conseguenza degli assiomi della logica induttiva: da qui il paradosso.

A prima vista il paradosso della conferma sembra destinato a infrangere ogni sforzo volto a costruire la logica induttiva intesa come analisi razionale delle inferenze non dimostrative. Tuttavia lo studio approfondito che ne è stato fatto ha reso un grande servizio a questa disciplina poiché ha reso i suoi cultori molto piú consapevoli di quanto vanno facendo aumentando pertanto l'efficacia delle loro ricerche. E vediamone le ragioni.

Appena si scopre un paradosso subito si cerca la sua « soluzione » e immaginiamo che anche il lettore sarà curioso di sapere quale « soluzione » è stata data al paradosso della conferma, forse per confrontarla con quella che ha trovato per conto suo. Come insegna però la storia della logica, ogni paradosso trova sempre piú di una soluzione e in una certa misura non ne trova mai perché ogni « soluzione » affronta il paradosso secondo un ben determinato punto di vista. Nel nostro caso purtroppo non possiamo presentare le « soluzioni » finora suggerite, in primo luogo, perché gli studi a esse legati sono ben lungi dall'essere conclusi; in secondo luogo, perché la parziali soluzioni finora trovate presuppongono un apparato tecnico che è fuor di luogo ora presentare. Una direzione di ricerca sembra però un fatto ormai acquisito e può essere esposta senza fare ricorso ad alcun tecnicismo. Si è cioè scoperto che il paradosso della conferma trova origine nell'aver messo sullo stesso piano le due famiglie di cui fanno parte rispettivamente le proprietà « essere una pietra di Marte » e « essere sferico ». Il paradosso potrà quindi risolversi se questa parità sarà lasciata cadere; ciò significa che se si vuole evitare la situazione paradossale di cui si è detto, si deve ammettere che le due famiglie « pesino » in modo diverso sulla determinazione della probabilità dell'ipotesi delle pietre di Marte. La incompletezza delle soluzioni finora trovate dipende dal fatto che non si è ancora in grado di dare un significato ai pesi che si assegnano alle famiglie.

Conclusione

Non ci sembra il caso di continuare in questa direzione, esponendo cioè le proposte di interpretazione dei «pesi» suddetti, quanto piuttosto di riflettere brevemente sulla influenza del paradosso sulla logica induttiva. Per far ciò è bene ricordare la situazione esistente prima della scoperta del paradosso. Non si tratta beninteso di un prima cronologico poiché il paradosso è una scoperta anteriore ai primi tentativi di costruire un sistema assiomatico per la logica induttiva, bensí di un prima metodologico. Nei primi tentativi che si sono compiuti per analizzare i metodi di assegnazione delle probabilità alle ipotesi induttive, era sembrato naturale assumere che tutte le famiglie dovessero essere trattate alla pari. Questa convinzione era suffragata dal fatto che la suddetta parità viene largamente usata nella valutazione statistica dei risultati sperimentali, come abbiamo mostrato quando abbiamo illustrato gli assiomi della logica induttiva e come mostra una breve riflessione sui metodi statistici.

La parità delle famiglie nei confronti della valutazione delle probabilità era quindi stata stabilita assiomaticamente. Convinti che la sua validità fosse universale, la si era applicata anche al caso delle probabilità delle leggi di natura (tale è infatti l'ipotesi delle pietre di Marte). La scoperta del paradosso segnò il fallimento del tentativo di affermare la validità universale dell'assioma che stabiliva la parità delle famiglie.

La conclusione che i logici induttivi hanno tratto da questa sconfitta è, come abbiamo detto, un arricchimento concettuale della loro disciplina. Ci si è cioè accorti che l'assioma dell'invarianza in questione può essere accettato solo in casi particolari, per esempio a livello delle situazioni sperimentali affrontate dalla statistica, mentre in altri casi, per esempio per le leggi di natura, deve essere sostituito con assiomi piú articolati che, stabilendo una gerarchia fra le famiglie, evitino di appiattirle tutte allo stesso livello.

È evidente che questa duttilità nell'uso degli assiomi è possibile solo se si pensa alla logica induttiva in modo non dogmatico, cioè solo se i suoi principi di base non sono intesi come verità incrollabili, bensí come strumenti che debbono essere accettati solo fin quando servono e che comunque devono confrontarsi dialetticamente con l'avanzare della ricerca. Ed è questa infatti la posizione dei moderni logici induttivi; essi sono in altri termini disposti ad accettare l'assioma dell'invarianza relativamente alle famiglie cosí come gli altri assiomi fin dove essi possono servire e sono invece pronti ad abbandonarli non appena si rivelassero dannosi per il progresso delle loro ricerche.

Abbiamo visto come il paradosso della conferma abbia aumentato la consapevolezza degli studiosi di logica induttiva e come lungi dal segnare la fine di questa disciplina le abbia permesso di scoprire nuove insospettate caratteristiche delle modalità d'assegnazione delle probabilità alle ipotesi induttive. La breve storia che abbiamo fatto di questo paradosso ci ha inoltre fornito l'opportunità di porre in risalto la posizione di fondo su cui si reggono le ricerche di logica induttiva, posizione che si ispira alla grande tradizione della moderna epistemologia di cui accetta le istanze filosofiche che sono poi le istanze su cui si fonda tutta la ricerca scientifica moderna.

Le molecole interstellari

Sono stati scoperti fino a oggi 26 tipi di molecole nel gas che si trova tra le stelle della nostra galassia. Tra esse vi sono ossido di carbonio, acqua, ammoniaca, acido solforico, formaldeide e alcool metilico

di Barry E. Turner

Tra le stelle della nostra galassia si trovano estese nuvole di gas e di polvere. Queste nuvole interstellari furono scoperte all'incirca 200 anni fa da William Herschel, che le descrisse come « buchi nel cielo » perché oscuravano la luce delle stelle situate dietro di esse. Per un lungo periodo la proprietà di oscurare fu attribuita alla polvere: minuscole particelle di composizione sconosciuta che arrossano e polarizzano la luce proveniente dalle stelle. Nel corso degli ultimi 35 anni, si è scoperto a poco a poco che il mezzo interstellare è composto anche di molte diverse specie di molecole libere, alcune delle quali alquanto complesse. Le molecole sono piú dense là dove la polvere è densa. Si trovano nelle regioni in cui sembra ci siano stelle in formazione e nell'atmosfera esterna delle stelle fredde. L'osservazione delle molecole fornisce informazioni sulle condizioni fisiche di queste regioni, che fino a poco tempo fa erano inaccessibili.

Lo studio quantitativo del mezzo interplanetario iniziò dopo che la spettrografia rese possibile l'analisi dettagliata della luce proveniente dalle stelle. Nel 1904, il tedesco Franz Hartmann suggerí che la riga di assorbimento del calcio ionizzato, visibile nello spettro di alcune stelle brillanti, avesse la sua origine nello spazio interstellare, cioè, in altre parole, che degli ioni di calcio (atomi privati di uno o piú elettroni) fossero presenti tra la Terra e le stelle e assorbissero la luce stellare a certe lunghezze d'onda. In seguito si scoprí che anche il sodio neutro, cioè non ionizzato, era un costituente del mezzo interstellare. Nel 1937 si scoprí che l'idrogeno è l'elemento piú abbondante nell'universo e che perciò doveva costituire la maggior parte del mezzo interstellare. Si riteneva che l'idrogeno fosse probabilmente presente, piú come atomo singolo che come molecola biatomica (H_2). Oggi sappiamo che, rispetto all'idrogeno, calcio e sodio sono presenti in quantità irrilevanti.

Nella regione visibile dello spettro, si può osservare idrogeno atomico (H) solo nello stato ionizzato, attraverso la emissione delle righe note come righe di ricombinazione di Balmer. Queste righe spettrali si producono quando un elettrone libero è catturato da un nucleo di idrogeno e scende verso i piú bassi livelli energetici dell'atomo. L'idrogeno ionizzato si trova nelle vicinanze delle stelle molto calde, dove è visibile sotto forma di nuvole brillanti dette nebulose di emissione o regioni H II. Per molti anni si è studiato l'idrogeno ionizzato di queste nebulose: a esso sono associate quantità molto piú piccole di elio atomico ionizzato, di ossigeno, di azoto, di carbonio e di tracce di altri elementi. Gli ioni sono indicatori della temperatura e della densità delle nuvole di gas che circondano le stelle piú calde, nuvole da cui evidentemente queste stelle si formano. Si scoprí molto tempo fa che le elevate temperature (10 000 °K) e le densità relativamente alte (100 atomi al centimetro cubo) non avevano alcun legame con le condizioni fisiche che prevalgono nelle regioni di spazio scure e fredde lontano dalle stelle calde.

La prima molecola interstellare fu scoperta nel 1937. Era il radicale chimico libero di carbonio e idrogeno (CH). Il radicale ionizzato (CH^+) e il radicale cianogeno (CN) furono identificati nel corso dei quattro anni seguenti negli spettri di alcune stelle brillanti blu-bianche di Tipo O e di Tipo B: spesso si trattava delle stesse stelle davanti alle quali si erano osservate le nuvole di calcio e sodio. Le tracce di questi costituenti non sono tuttavia utili per lo studio del mezzo interstellare in generale. Una ragione di questo fatto è che si possono osservare solo davanti alle stelle piú brillanti. Un'altra è che le nuvole che le contengono devono essere sufficientemente dense da produrre una riga di assorbimento apprezzabile, ma non troppo dense da far sí che la luce che proviene dalla stella non sia troppo attenuata. Sono perciò necessarie condizioni molto particolari. In terzo luogo, le osservazioni a lunghezze d'onda visibile possono penetrare il mezzo interstellare solo per 2000 o 3000 anni luce: infatti queste lunghezze d'onda sono notevolmente assorbite dalla polvere interstellare. In quarto luogo, l'osservazione di tracce di componenti quali calcio, sodio, CH, CH^+, e CN non danno alcuna informazione sulla quantità totale di gas presente tra le stelle. Per di piú, si può provare che non sono collegati fisicamente con le poche stelle molto calde in direzione delle quali sono osservati.

Non si sarebbe mai potuti uscire da questo vicolo cieco, se non vi fosse stato il rapido sviluppo della radioastronomia nel corso degli ultimi 20 anni. Infatti, uno dei maggiori trionfi della radioastronomia fu la scoperta, nel 1951, di idrogeno atomico interstellare per mezzo della riga spettrale alla lunghezza d'onda radio di 21 centimetri. Si comprese in fretta che le osservazioni nel campo radio potevano penetrare

La Grande Nebulosa di Orione è una nuvola di polvere e di gas, nella quale vi sono stelle in formazione. Immerse nella nebulosa ci sono quattro stelle brillanti giovani e molto calde, note come Gruppo del Trapezio. Esse costituiscono la sorgente di energia della nebulosa e la fanno irraggiare a molte lunghezze d'onda. Il colore verdastro della nebulosa dipende dal fatto che è stata eseguita a una lunghezza d'onda di 5007 angstrom.

completamente attraverso la Galassia, che ha un diametro di circa 60 000 anni luce, perché le radioonde non sono assorbite in maniera apprezzabile dalla polvere interstellare. Negli anni dopo il 1951, la riga dell'idrogeno a 21 centimetri venne immediatamente usata come strumento per studiare il gas da un'estremità all'altra della nostra galassia, gas che costituisce tra il 5 e il 7 per cento della massa totale della Galassia. Nel corso delle esplorazioni fatte alla riga di 21 centimetri, si tracciò l'intera struttura a spirale della Galassia; prima la si era appena intuita attraverso la distribuzione delle stelle.

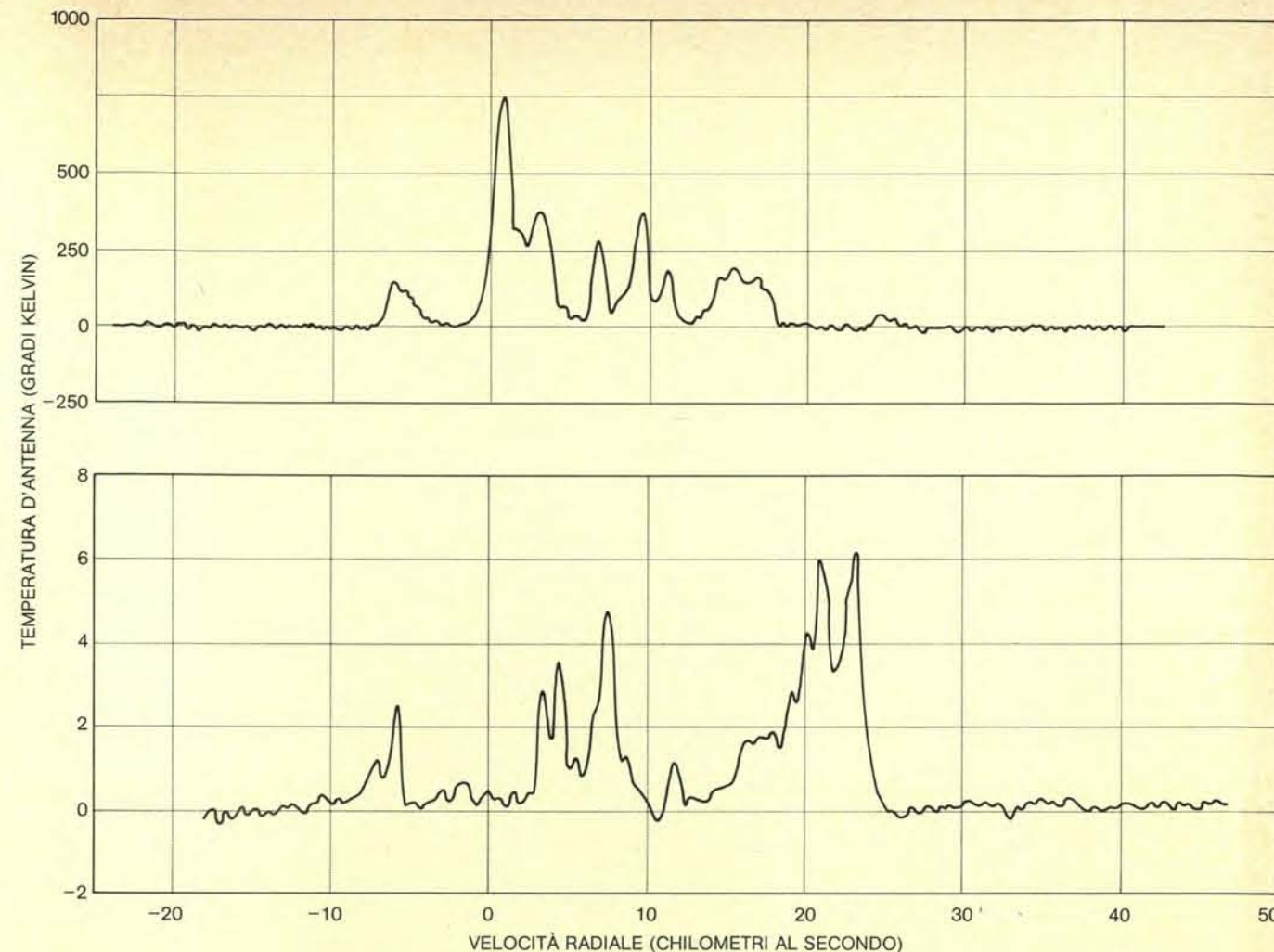
Dall'esame della riga a 21 centimetri dell'idrogeno, che tipo di considerazioni fisiche e chimiche si potevano dedurre sul gas interstellare? Le prime immagini non erano accurate, ma indicavano che il gas aveva in media una temperatura di 100 °K e una densità compresa tra un atomo e dieci atomi al centimetro cubo nelle braccia a spirale della Galassia e una densità di 0,1 atomo al centimetro cubo nelle zone tra le braccia. Perfezionamenti di tipo sperimentale e considerazioni teoriche hanno fatto elaborare un modello a « due componenti » del mezzo interstellare. Una componente è formata da nuvole fredde di gas con una densità compresa tra 10 e 100 atomi al centimetro cubo e una temperatura pari o inferiore ai 100 °K. Queste nuvole si trovano in equilibrio di pressione con la seconda componente, un gas caldo con una densità di circa 0,1 atomo al centimetro cubo e una temperatura che forse raggiunge i 10 000 °K. Il gas all'interno delle braccia a spirale è composto soprattutto dalla componente calda. Piú o meno era questa l'immagine che si aveva del mezzo interstellare nel 1968, anno che vede la nascita di quella che oggi è nota come astronomia molecolare. Nel 1963 i radio astronomi avevano scoperto il radicale ossidrilico (ossidrile, OH). Nel 1968 l'ossidrile era stato scoperto in varie direzioni della Galassia: quasi tutte si trovavano in direzione di regioni H II e di stelle giovani, rivelabili soprattutto nel campo dell'infrarosso. Poiché le regioni H II irraggiano a una temperatura piú alta della temperatura del mezzo interstellare e del gas OH che vi è immerso, gli osservatori si aspettavano che il gas OH avrebbe in parte assorbito la radiazione proveniente dalle regioni H II. Perciò ci si aspettava che la riga spettrale del gas OH sarebbe stata vista in assorbimento nello spettro radio delle regioni H II.

L'emissione dell'ossidrile

In molti casi, tuttavia, l'OH era osservato non in assorbimento, ma come una stretta e intensa riga di emissione nello spettro radio delle regioni H II (*si veda l'illustrazione nella pagina a fronte*). Se, come ci si aspettava, le molecole OH avessero assorbito la radiazione, ciò avrebbe indicato che la maggior parte di esse si trovava a un livello energetico basso. Il comportamento opposto indica che i livelli energetici delle molecole OH sono piuttosto differenti da quelli che si avrebbe se il gas OH fosse in equilibrio con l'ambiente freddo circostante. Evidentemente il gas OH è « pompato » da qualche meccanismo in uno stato eccitato in cui è capace di amplificare la radiazione di fondo. Questo meccanismo corrisponde a un maser interstellare.

Il maser è il precursore del laser e amplifica le radioonde invece della luce visibile. Nel maser interstellare le molecole di OH sono eccitate, per cui la maggior parte di esse si trovano in un livello energetico alto anziché basso. Quando una delle molecole cade in un livello energetico basso emette un quanto di energia a una certa riga dello spettro radio; l'energia emessa stimola altre molecole a comportarsi nello stesso modo e qualche sorgente di energia nella nuvola interstellare continua a riportare le molecole nello stato eccitato cosicché il processo continua.

Le osservazioni della strana emissione OH, però, anziché contribuire a dare una visione nuova delle condizioni fisiche del mezzo interstellare, ponevano una quantità di nuovi problemi sulla eccitazione delle molecole del mezzo. Nella vicinanza delle zone a forte emissione OH il mezzo è molto denso rispetto ai parametri di densità interstellare, addirittura 10^8 particelle al centimetro cubo. Queste zone sembrano essere confinate all'atmosfera di stelle fredde e giovani o protostelle, oggetti che stanno diventando stelle.

Le curve di livello sovrapposte su un negativo della Nebulosa di Orione mostrano la distribuzione delle molecole interstellari di formaldeide (H_2CO), alla lunghezza d'onda di due millimetri. I numeri che si trovano sulle linee di livello sono le temperature di brillanza, una misura, espressa in gradi Kelvin, dell'intensità della radiazione emessa al di sopra della radiazione cosmica di fondo di 3 °K. Le due ragioni di piú forte emissione molecolare non corrispondono a nessun oggetto osservato nel visibile ma coincidono con zone di forte emissione infrarossa. Uno di questi oggetti è la nebulosa infrarossa scoperta da D. E. Kleinmann e F. J. Low dell'emissione maser da radicale ossidrilico (OH) e vapor acqueo (H_2O). Questa sorgente, il cui spettro è mostrato nella pagina a fronte, si trova vicino a una sorgente brillante simile a una stella (*C*) che emette radiazione infrarossa. Essa fu scoperta da E. E. Beckil e G. Neugebauer dell'Istituto di Tecnologia della California. Si ritiene che queste nuvole molecolari si trovino dietro alla brillante Nebulosa di Orione, in una regione densa di gas non ancora ionizzato delle stelle centrali del Trapezio. In questa regione, il solfuro di carbonio (CS) si trova distribuito in gran parte nello stesso modo della formaldeide; nella stessa regione si trova una nuvola di acido cianidrico (HCN) che ha dimensioni circa doppie delle nuvole di solfuro di carbonio e formaldeide, ed è simile a una nuvola di una molecola X-ogena non identificata. Si osserva ossido di carbonio (CO) in una nuvola che ha un diametro di un grado (20 volte quello della nuvola di solfuro di carbonio).

Lo spettro dell'acqua (*in alto*) e del radicale ossidrilico (*in basso*) mostrano che queste molecole interstellari emettono piú radiazione di quanta ne assorbono. Gli spettri di emissione sono stati ottenuti nella direzione di oggetti infrarossi della Nebulosa di Orione. Sono calibrati sulla temperatura dell'antenna ricevente la radiazione (*scala verticale*); la temperatura di brillanza è uguale al prodotto della temperatura di antenna per il quadrato del rapporto tra la dimensione del fascio del telescopio e la dimensione angolare della sorgente. Misure interferometriche che hanno stabilito che la dimensione angolare di queste due sorgenti di emissione non supera i 0,005 secondi d'arco. Questo implica che la temperatura di brillanza è in eccesso di 10^{13} °K sia per l'acqua sia per il radicale ossidrilico. L'emissione delle molecole d'acqua è altamente variabile e cambia in una scala di tempo di pochi giorni; l'emissione dell'ossidrile è altamente polarizzata. Queste caratteristiche indicano che funziona un maser interstellare che « pompa » le molecole in uno stato eccitato: in questo modo esse amplificano la debole radiazione locale.

Le molecole piú complesse

Le molecole interstellari composte di piú di due atomi furono scoperte per la prima volta alla fine del 1968. Esse hanno profondamente alterato il nostro concetto di chimica interstellare e hanno modificato le nostre visioni sulle condizioni chimiche dello spazio. Nel 1968 la maggior parte degli astronomi riteneva che la densità del mezzo interstellare fosse cosí bassa da rendere difficile la combinazione di piú di due atomi. Essi prevedevano di trovare nello spazio interstellare solamente molecole biatomiche, e che anche queste avessero una breve vita a causa degli effetti distruttivi della radiazione ultravioletta e dei raggi cosmici. Queste idee dovettero essere drasticamente riviste quando un gruppo dell'Università della California a Berkeley (Charles H. Townes, William J. Welch, A. C. Cheung, David M. Rank e D. D. Thornton) trovò l'ammoniaca (NH_3) in numerose nuvole interstellari, nella zona del centro galattico. Poco dopo il gruppo di Berkeley scoprí segni di emissione da vapor acqueo in parecchie regioni della Galassia.

Anche se in seguito si ebbero a scoprire molecole piú complesse, l'acqua e l'ammoniaca avevano ormai stabilito due importantissime realtà sulle nuvole molecolari interstellari. 1) Le nuvole sono piuttosto dense rispetto a tutte le altre regioni conosciute dello spazio interstellare; 2) le condizioni fisiche che eccitano le molecole in modo da farle irraggiare o assorbire radiazione, sono differenti dalle condizioni terrestri. In alcuni casi le condizioni fisiche danno luogo a un maser interstellare; in altri creano quello che si potrebbe chiamare un frigorifero interstellare. Una molecola come la formaldeide (H_2CO) assorbe infatti piú energia di quanto sembra permesso dalle condizioni ambientali, almeno se si devono ritenere valide le leggi della termodinamica terrestre.

Dove si trovano le molecole

Dal 1968, la velocità con cui nuove molecole interstellari sono state scoperte è cresciuta esponenzialmente. L'elenco arriva oggi a 26 molecole. Dove si trovano queste molecole? Come si formano e come si distruggono? Come può l'astronomo utilizzare i loro segnali per capire i processi fisici e chimici che hanno luogo nello spazio interstellare?

Nel corso degli ultimi due anni la Galassia è stata intensamente osservata alle lunghezze d'onda caratteristiche del radicale OH, della formaldeide e dell'ossido di carbonio (CO). Le osservazioni hanno mostrato che queste molecole, come l'idrogeno atomico, sono fortemente concentrate in direzione del piano centrale della Galassia in uno strato che, in prossimità del Sole ha uno spessore inferiore ai 1000 anni luce. Le molecole sembrano distribuite uniformemente nel disco e raggiungono concentrazioni maggiori nelle vicinanze del centro della Galassia. L'ossido di carbonio e la formaldeide sono distribuiti allo stesso modo mentre la maggior parte delle altre molecole interstellari si osserva solo in pochissime regioni. Non si sa ancora se la distribuzione osservata sia dovuta al fatto che esse sono assenti, oppure al fatto che sono semplicemente in uno stato non eccitato per cui non emettono né assorbono segnali misurabili.

È oggi chiaro che le molecole si trovano nello spazio interstellare nelle regioni che hanno una alta densità di particelle. Ci sono due ragioni per questa correlazione. La prima è che quanto più è alta la densità del mezzo interstellare, tanto maggiore è la possibilità che si formino molecole dagli atomi presenti. Le molecole si formano o per semplice collisione tra gli atomi del gas interstellare, o per una loro interazione sulla superficie di una particella di polvere interstellare. La seconda ragione è che, delle molecole interstellari conosciute, molte devono essere eccitate a determinati livelli energetici per essere in grado di emettere quantità apprezzabili di radiazione. Questi livelli si generano soprattutto attraverso collisioni continue con altri atomi o molecole del gas: quanto più è denso il gas, tanto più frequentemente questo accade. Ma quali sono le regioni interstellari che possiedono questa alta densità?

Nuvole scure e protostelle

Entro una distanza di 1000 anni luce dal Sole, ci sono circa una dozzina di nuvole scure, che nelle fotografie appaiono come buchi sullo sfondo di stelle; probabilmente nella Galassia esistono circa 3000 nuvole di tipo analogo. Hanno di solito un diametro di 12 anni luce e la massa della polvere che vi si trova è circa 20 volte la massa del Sole. Alcune molecole interstellari – per esempio l'ossidrile, l'ammoniaca, la formaldeide e l'ossido di carbonio – sono state individuate in queste nuvole, ma la maggior parte delle molecole non è stata trovata. Da queste osservazioni si è dedotto che nelle nuvole la massa totale del gas supera più di 100 volte la massa della polvere. La densità del gas è solitamente compresa tra poche centinaia di particelle e 10 000 particelle al centimetro cubo. La temperatura è bassa: non più di 4 °K in alcune nuvole e forse fino a 25 °K nelle nuvole più calde. Dalla dimensione e la massa delle nuvole scure si può calcolare che sono gravitazionalmente instabili e stanno collassando per dare origine a stelle.

Si ritiene che le nuvole scure si contraggano lentamente nei cosiddetti globuli: piccole nuvole di polvere che sulle fotografie delle nebulose di emissione si manifestano come macchioline completamente opache. Si conosce poco del contenuto gassoso di questi oggetti, e stime della loro massa, danno solo un limite inferiore. Tuttavia, recentemente, in alcuni globuli sono stati scoperti l'ossidrile e la formaldeide. Analisi preliminari indicano che la temperatura dei globuli è ancora più bassa di quella delle nuvole di polvere più estese e che la densità è un po' più alta.

Si ritiene che un globulo sia uno degli stadi finali nel collasso da una nuvola interstellare a una protostella, una concentrazione più alta di polvere e gas, con all'incirca le dimensioni del sistema solare, che sta per diventare una vera stella. Nello stadio protostellare del collasso, la densità e la temperatura vicino al centro dell'oggetto si alza a sufficienza per provocare l'irraggiamento piuttosto intenso nella regione infrarossa dello spettro. Il lavoro teorico di R. B. Larson dell'Università di Yale, mostra che la protostella collassa molto più velocemente vicino al centro che nelle regioni esterne: così le stelle appena nate dovrebbero essere circondate da un esteso inviluppo di polvere e di gas. A sua volta l'inviluppo sarebbe circondato da quello che rimane dalla nuvola interstellare. Se la nuova stella è molto luminosa di Tipo O o di Tipo B, dovrebbe emettere nell'ultravioletto una quantità di radiazione sufficiente per ionizzare l'idrogeno dalla nuvola circostante e formare una regione H II. Questa a sua volta dovrebbe essere circondata da una nuvola di gas non ionizzato, che non sarebbe stato raggiunto dalla radiazione ionizzante.

In che modo le molecole si inquadrano in questa immagine? Esse sembrano associate a ogni stadio dal processo di contrazione di una nuvola interstellare. Abbiamo visto che l'ossidrile, la formaldeide e l'ossido di carbonio si osservano in tutte le regioni delle braccia a spirale e forse anche tra le braccia. Le nuvole di polvere più estese contengono anche ammoniaca. Le altre molecole si osservano solo quando ci troviamo in regioni in cui si ritiene che vi siano delle stelle in formazione, regioni in cui le molecole possono essere associate alle protostelle. Oppure possono esistere nelle nuvole molto massicce, dense e non ionizzate, che circondano le regioni H II formatesi da stelle appena nate.

Un buon esempio del secondo tipo è costituito dalla Grande nebulosa di Orione. Immerso nella Nebulosa di Orione c'è un gruppo di stelle giovani molto calde, noto come Gruppo del Trapezio, che sono la fonte di energia della nebulosa. Attorno a essa vi è una estesa nuvola di idrogeno neutro; appena dietro c'è una densa nuvola scura nel cui centro gli astronomi dell'infrarosso hanno scoperto alcuni oggetti che sembrano protostelle.

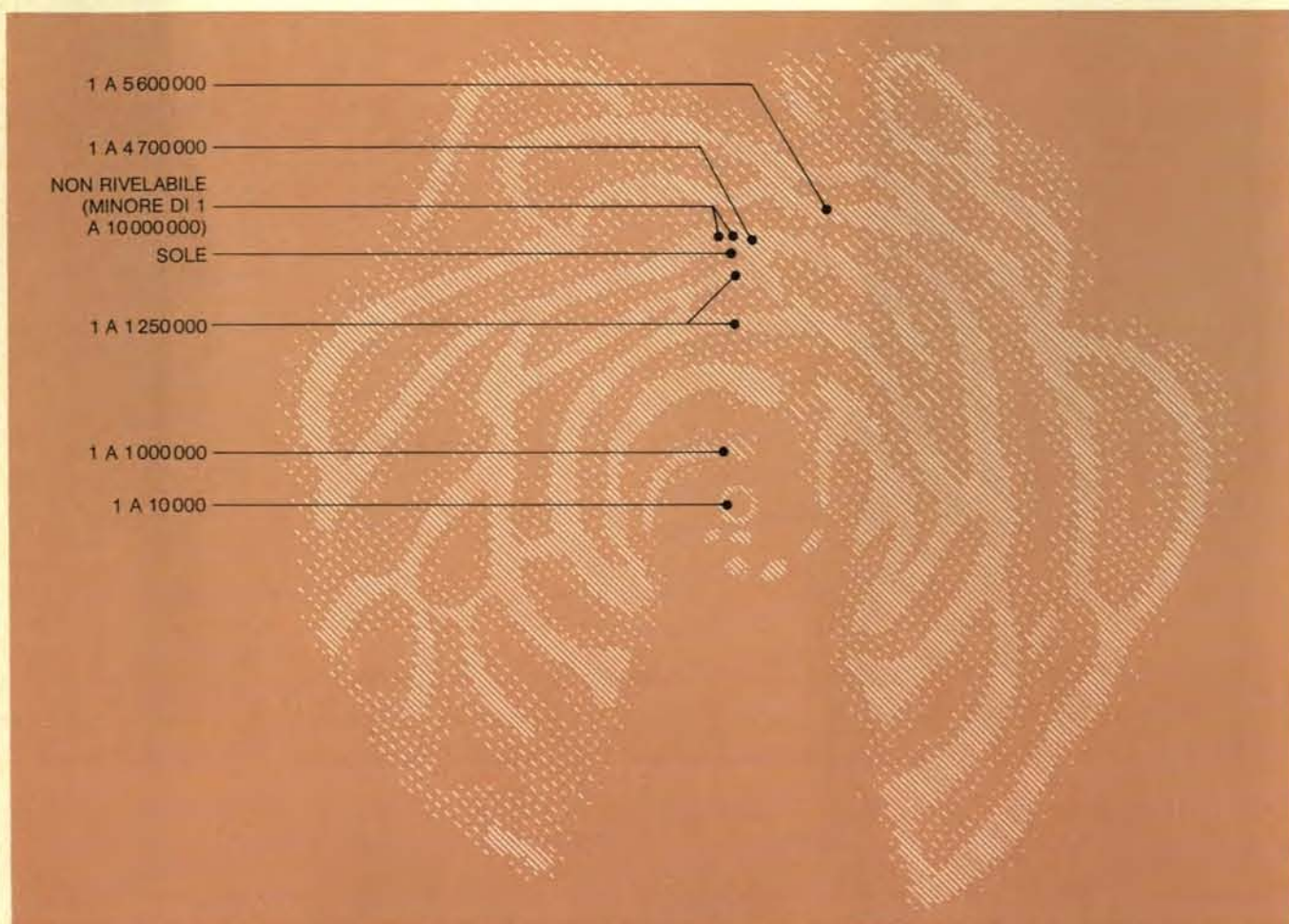
Si trovano molecole in tutta la regione della Nebulosa di Orione. Una grande nuvola di ossido di carbonio esce dalla nebulosa e entra nella nuvola circostante di idrogeno neutro: occupa circa un grado d'arco nel cielo e ha una massa gassosa pari a circa 100 masse solari. Il radicale ossidrilico ha una distribuzione simile. Nuvole più piccole di acido cianidrico (HCN) e di molecole non identificate riconoscibili come radicali X-ogeni, si vedono proiettate sulla parte centrale della nebulosa. Nelle immediate vicinanze degli oggetti infrarossi, si osservano forti concentrazioni di formaldeide, alcool metilico (CH_3OH), solfuro di carbonio (CS), cianogeno (CN), ammoniaca e cianoacetilene (HC_3N). Ci sono anche nuvole molto piccole di radicale ossidrilico e vapore acqueo: sono fortemente eccitate ed emettono intensi segnali radio di tipo maser.

La distribuzione di queste molecole nella Nebulosa di Orione non è casuale. Le molecole possono esistere in quantità osservabile solo in regioni che soddisfano due condizioni: 1) la velocità con cui le molecole si formano è maggiore di quella con cui si distruggono: 2) le molecole si trovano in una regione che può essere eccitata ai livelli energetici che si osservano. Le regioni ad alta densità sono buone candidate a soddisfare queste richieste. Il radicale ossidrilico (quando si osserva in assorbimento) e l'ossido di carbonio si osservano in livelli bassi di eccitazione e non vengono facilmente distrutti dalla luce. Questo spiega perché si osservano in tutta la galassia e perché occupano una regione maggiore della Nebulosa di Orione delle altre molecole. Le molecole che si trovano in una piccola zona centrale della Nebulosa di Orione, non si osservano comunemente nella Galassia, ma solo in zone apparentemente molto dense che sono vicine a regioni H II. In realtà delle 26 molecole interstellari che si conoscono, 8 sono state viste in un'altra regione H II vicino al centro galattico, e non nella Nebulosa di Orione. Questa regione, l'unica in cui queste molecole sono state viste, è nota come Sagittario B2. La caratteristica particolare del Sagittario B2 non è la temperatura (circa 100 °K) ma l'enorme densità (fino a 10^8 particelle al centimetro cubo) e le sue dimensioni (circa 20 anni luce): questo fa sì che, nella linea della visuale, la nebulosa abbia una densità 100 e più volte maggiore di qualsiasi altra nuvola interstellare.

Maser interstellari

Una delle prime indicazioni del fatto che le molecole interstellari erano probabilmente associate a protostelle, venne dall'emissione maser del radicale ossidrilico e del vapore acqueo. A differenza del radicale ossidrilico, che si osserva spesso come normale riga di assorbimento nello spettro radio continuo delle sorgenti, il vapore acqueo si vede solo come potente riga di emissione. Per di più, si osserva solo in regioni che sono forse più piccole del sistema solare e forse anche più piccole dell'orbita della Terra. Le righe di emissione del vapore acqueo sono talmente intense, che se fossero irraggiate da un corpo riscaldato nella maniera usuale, la temperatura del corpo dovrebbe essere dell'ordine di 10^{13} °K. Simili temperature possono essere collegate alla energia cinetica delle molecole o con qualsiasi temperatura di radiazione solo se la radiazione è stata amplificata. Queste caratteristiche implicano l'esistenza di un meccanismo di maser interstellare.

Subito dopo che nel 1968 furono scoperti maser interstellari ad acqua, si scoprì che si trovano sempre in regioni in cui l'ossidrile emette una potente radiazione maser. La temperatura di radiazione equivalente per il radicale ossidrilico è circa la stessa che per l'acqua. In alcune di queste sorgenti, la quantità di energia irraggiata nelle due strette righe spettrali dell'ossidrile e dell'acqua è pari all'energia emessa dal Sole in tutto lo spettro. L'energia deve indirettamente venire dal collasso gravitazionale della protostella; l'emissione provoca il raffreddamento della nuvola e la fa collassare più in fretta.

Le teorie che cercano di spiegare una simile eccitazione dell'ossidrile e dell'acqua, e la ragione per cui si tro-

Questa « pianta » della Galassia mostra la distribuzione delle molecole di ossidrile nel mezzo interstellare. La Galassia è stata osservata alla lunghezza d'onda radio della riga spettrale dell'atomo d'idrogeno; la mappa mostra come è distribuito l'idrogeno nelle braccia a spirale. Il rapporto tra il radicale ossidrilico e l'idrogeno cresce di un fattore mille tra la regione vicino al sistema solare e il centro galattico. Qui si trova un radicale ossidrilico ogni 10 000 atomi di idrogeno, ma osservazioni in altre direzioni rivelano meno di un radicale ossidrilico ogni quattro milioni di atomi d'idrogeno o non rivelano l'assorbimento.

vano insieme, considerano vari tipi di pompaggio e gli effetti della forte radiazione infrarossa. Questi processi sono efficaci solo se le densità raggiungono 10^8 particelle al centimetro cubo e le temperature vanno da alcune centinaia a 1000 °K. Queste condizioni esistono solo in protostelle quali le sorgenti infrarosse nella Nebulosa di Orione. In realtà, insieme con alte densità, si richiede una considerevole quantità di energia (pari a 700 °K) per eccitare le molecole d'acqua e farle irraggiare nella maniera osservata. Questo esclude che si possa osservare acqua interstellare in oggetti che non hanno queste particolari condizioni.

Le condizioni che si richiedono per generare l'azione maser dell'ossidrile e dell'acqua, esistono anche nelle regioni esterne di certe stelle molto fredde. Questo ci porta all'ultima categoria di oggetti celesti in cui si osservano delle molecole. Un certo numero di molecole – idrogeno biatomico (H_2), acqua, ossido di carbonio, cianogeno, carbonio biatomico (C_2), metilidina (CH), acido cianidrico (HCN) e acetilene (C_2H_2) – sono note da molto tempo per i loro spettri ottici e infrarossi provenienti dall'atmosfera di stelle rosse fredde Tipo M e delle stelle note come variabili Mira Ceti. Tuttavia, solamente nel 1968, W. J. Wilson e Alan H. Barret dell'Istituto di Tecnologia del Massachusetts, scoprirono forti emissioni del radicale ossidrilico alle lunghezze d'onda radio, provenienti da stelle infrarosse, non associate a regioni H II. Di queste stelle, molte sembrano giovanissime; alcune di queste sono le variabili Mira Ceti: precedentemente G. Neugebauer e Robert B. Leighton dell'Istituto di Tecnologia della California, avevano scoperto che esse erano fonti sorgenti di radiazione infrarossa. Due anni dopo, anche l'emissione maser da vapore acqueo fu scoperta nelle variabili Mira Ceti, si ritiene che i maser siano pompati dalla radiazione infrarossa.

Marvin M. Litvak del MIT, ha mostrato che il radicale ossidrilico, sotto l'azione di un pompaggio infrarosso, emetterebbe radiazioni alla frequenza di 1612 megahertz (10^6 cicli al secondo). Nelle stelle infrarosse, l'OH irraggia in prevalenza a questa frequenza. Questo invece non è il caso del radicale ossidrilico nelle protostelle, che irraggia in prevalenza a 1665 megahertz e può essere pompato da meccanismi chimici. Si ritiene che l'emissione dell'ossidrile e dell'acqua abbia origine dal guscio di gas e di polvere che circonda queste stelle. L'osservazione del guscio ha rivelato uno spostamento Doppler delle righe dello spettro che suggerisce che esso si espanda o si contragga a una velocità tra i 10 e i 30 chilometri al secondo. Quasi tutte le stelle nella cui atmosfera esterna si trovano ossidrile e vapore acqueo, sembrano contenere piú ossigeno che carbonio. In altri tipi di stelle fredde è vero l'opposto. Nelle stelle « di carbonio », non si trovano l'ossidrile e il vapor acqueo mentre sono state osservate nella banda radio e infrarossa molecole contenenti carbonio, come ossido di carbonio, solfuro di carbonio (CS), HCN e CN. Le osservazioni radio si sono rivelate particolarmente interessanti perché indicano in queste molecole la presenza di carbonio-12, l'isotopo piú comune del carbonio, ma anche di carbonio-13, isotopo piú raro. In queste stelle il rapporto tra carbonio-12 e carbonio-13 è solo di 4 : 1 circa; sulla Terra e nella maggior parte del mezzo interstellare il rapporto è di 89 : 1.

Molecole come sonde interstellari

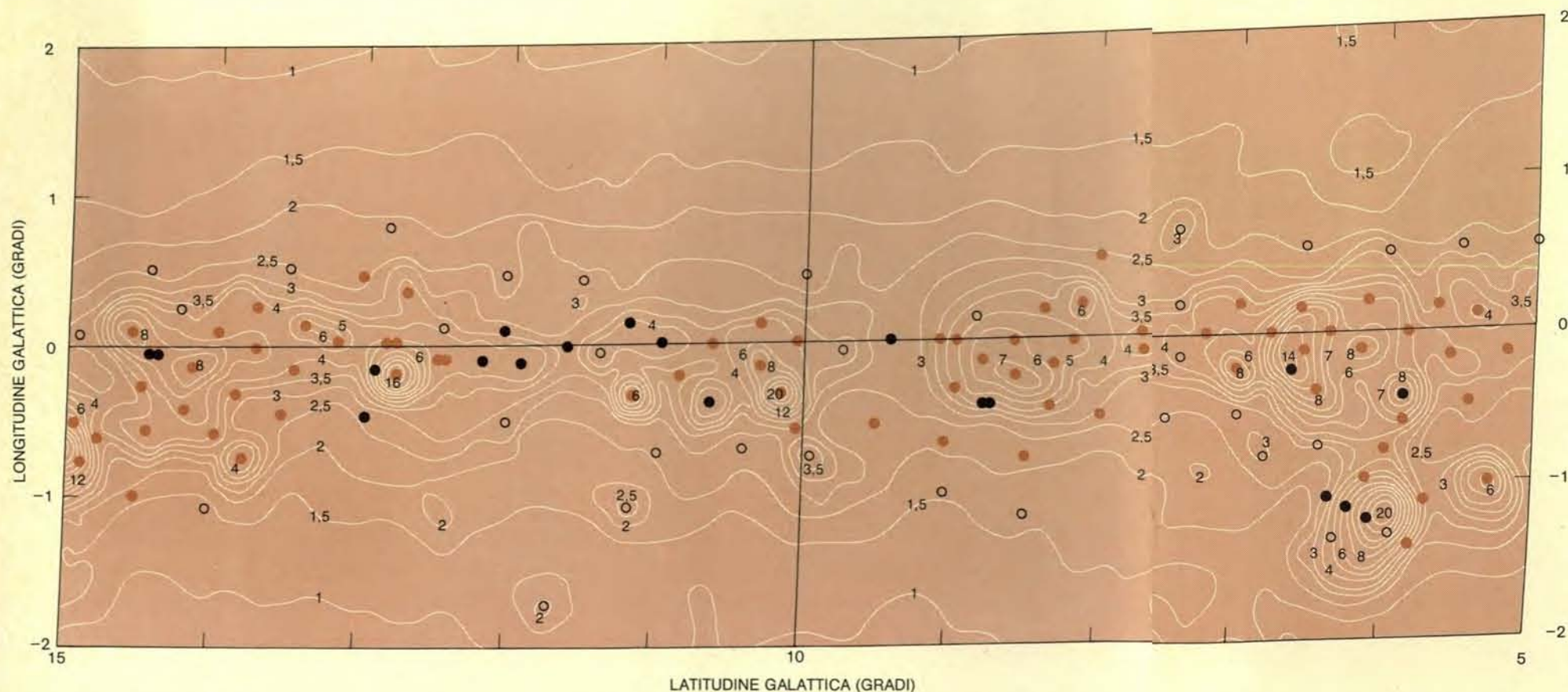
Scoprendo le molecole, misurando la loro posizione e lo spostamento Doppler delle righe spettrali, si possono ottenere informazioni dirette sulla costituzione, posizione e velocità delle nuvole interstellari. Se poi si fa un'analisi piú accurata dell'intensità delle righe e la si unisce ad altre informazioni, si possono avere notizie sulle condizioni fisiche e sulla dinamica all'interno delle nuvole. Le particolari righe spettrali che indicano la presenza di molecole, si formano quando esse, o gli elettroni che le formano, cambiano il loro stato di moto. Ogni molecola tende a ruotare attorno al proprio asse di simmetria. Cambiamenti nella rotazione la fanno irraggiare o assorbire energia elettromagnetica a lunghezze d'onda che si trovano normalmente nella banda delle microonde (lunghezze d'onda comprese tra un millimetro e 6 centimetri). Gli atomi della molecola vibrano e ruotano uno intorno all'altro. Anche il moto vibrazionale può cambiare, provocando l'irraggiamento o l'assorbimento, da parte della molecola, di radiazione infrarossa. Inoltre gli elettroni dei vari atomi possono passare da un'orbita all'altra: queste transizioni provocano una riga di emissione o di assorbimento nelle regioni visibili o ultraviolette dello spettro.

Il solfuro di carbonio è un esempio del piú semplice tipo di molecola: una molecola lineare diatomica. Nei livelli piú bassi di energia elettronica e vibrazionale, gli unici stati popolati nello spazio interstellare, i moti possibili della molecola sono il semplice movimento in una direzione e una rotazione continua. Le leggi della meccanica quantistica dicono che i corrispondenti livelli energetici formano una semplice « scala ». La molecola si può muovere su e giú sulla scala quando urta con un'altra molecola o un altro tipo di particella, o quando assorbe o emette un fotone (un quanto di radiazione). Lasciata a se stessa, la molecola emette spontaneamente un fotone e cade nel livello fondamentale, cioè il livello energetico piú basso, in un tempo definito. I livelli energetici sono definiti dai numeri quantici J, che sono una misura del momento rotazionale della molecola.

Il solfuro di carbonio impiega in media quattro ore per cadere dal primo livello energetico al livello energetico fondamentale. Il tempo che la molecola passa nei livelli energetici piú alti è minore. Gli astronomi rivelano la radiazione proveniente dai livelli energetici piú alti, e quindi ci deve essere qualche meccanismo che mantiene le molecole del solfuro di carbonio in questi livelli e che si oppone alla tendenza di decadere spontaneamente. Potrebbero agire dei campi di radiazione, ma dovrebbero essere molto piú intensi di quelli che si osservano direttamente in queste nuvole. Un meccanismo piú probabile di eccitazione è la collisione con altre particelle.

Per i tipi di particelle che consideriamo e le temperature a cui si trovano, possiamo stimare qual è la densità necessaria per mantenere negli stati eccitati molecole come il solfuro di carbonio. Si hanno normalmente densità di un milione di particelle al centimetro cubo per le nuvole di solfuro di carbonio che si trovano vicino alle regioni H II. Altre molecole che hanno livelli energetici simili al solfuro di carbonio sono l'ossido di carbonio, l'ossido di silicio (SiO), il solfuro carbonilico (OCS), il cianuro d'idrogeno e la cianoacetilene (HC_3N). L'ossido di carbonio decade spontaneamente molto piú lentamente delle altre molecole lineari. Perciò si richiede una quantità minore di eccitazione dovuta a collisione per mantenere i livelli eccitati. Questa è una delle ragioni per cui l'ossido di carbonio si osserva in regioni dello spazio piú estese di altre molecole e lo si trova anche in nuvole interstellari in cui la densità è di solo 100 particelle al centimetro cubo circa.

Osservando le differenti transizioni delle molecole all'interno di una nuvola, è possibile individuare zone di densità differente. I livelli energetici piú alti richiedono densità maggiori per provocare la dovuta eccitazione. Una riga spettrale ha una certa ampiezza che è determinata in parte dalla temperatura della stella (quanto piú la temperatura è alta, tanto piú larga è la riga), in parte alla turbolenza dovuta ai moti su larga scala entro la nuvola. Nella maggior parte dei casi, la turbolenza domina gli effetti di temperatura. In questi casi, tanto piú è stretta la riga spettrale, tanto piú è piccola la turbolenza e perciò tanto piú è piccola la regione della nuvola che emette la riga. Per esempio, la riga spettrale emessa dal solfuro di carbonio quando decade dal secondo livello energetico al primo (da $J = 2$ a $J = 1$), è emessa, a quanto si osserva, in due distinti intervalli di velocità (*si veda l'illustrazione a pag. 60 in basso*). La riga che si produce quando la molecola decade dal primo livello energetico al livello fondamentale (da $J = 1$ a $J = 0$) viene emessa su un intervallo di velocità molto maggiore. Questo fatto viene interpretato come una indicazione che, entro la grande nuvola di solfuro di carbonio, si trovano due regioni piú piccole di densità maggiore. L'osservazione di molecole piú eccitate dovrebbe raffinare maggiormente la mappa di densità.

Come si possono determinare le temperature di una nuvola da osservazioni molecolari? Si scopre che le molecole, come il solfuro di carbonio, che irraggiano rapidamente la loro energia, sono delle sonde di temperatura piuttosto inefficaci; altre molecole, come l'ossido di carbonio e l'ammoniaca, sono piú utili. Se una molecola di carbonio irraggia la sua energia molto piú lentamente di quanto non sia eccitata attraverso il processo di collisione, allora si mette in equilibrio termico con l'ambiente. In queste condizioni, l'intensità delle varie transizioni tra livelli energetici si può calcolare esattamente come funzione della sola temperatura.

Visione laterale della Galassia, in prossimità del centro galattico; la mappa alla lunghezza d'onda radio di 11 centimetri è dovuta a Wilhelm J. Altanhoff del Max Planck Institute di Bonn. Le linee di livello sono misure della intensità della radiazione proveniente da varie sorgenti radio simili alle regioni H II (regioni di gas ionizzato vicino a stelle giovani e calde) e i resti di esplosioni di supernove. La radiazione ha uno spettro continuo, diverso dalle « righe » spettrali che caratterizzano atomi e molecole, in cui si verifica emissione o assorbimento solo in intervalli molto stretti e determinati di lunghezza d'onda. I numeri sulle linee di contorno, sono temperature di brillanza. I punti rappresentano regioni in cui l'autore ha cercato righe di emissione o di assorbimento del radicale ossidrilico. I punti neri indicano zone in cui l'OH assorbe parte della radiazione alla lunghezza d'onda caratteristica di 18 centimetri. I cerchi indicano zone in cui l'OH non è stato osservato. Molte sorgenti di emissione OH sono anche sorgenti di forte emissione maser da vapor acqueo. Come si vede, l'OH è un componente molto diffuso del disco galattico ed è strettamente confinato in una regione vicino al piano galattico. L'ossido di carbonio e la formaldeide sono distribuiti nello stesso modo. La maggior parte delle altre molecole è stata osservata in poche sorgenti.

ANNO	MOLECOLA	FORMULA	LUNGHEZZA D'ONDA	NUMERO DI REGIONI	TELESCOPIO
1937	METILIDINA (IONIZZATA)	CH^+	3958 ANGSTROM*	88	100 POLLICI, MWO
1937	METILIDINA	CH	4300 ANGSTROM*	60	100 POLLICI, MWO
1939	CIANOGENO	CN	3875 ANGSTROM*	14	100 POLLICI, MWO
1963	OSSIDRILE	OH	18,0 CENTIMETRI*	~600	84 PIEDI, LL
1968	AMMONIACA	NH_3	1,3 CENTIMETRI*	12	20 PIEDI, HCO
1968	ACQUA	H_2O	1,3 CENTIMETRI*	35	20 PIEDI HCO
1969	FORMALDEIDE	H_2CO	6,2 CENTIMETRI*	~150	140 PIEDI, NRAO
1970	OSSIDO DI CARBONIO	CO	2,6 MILLIMETRI	60	36 PIEDI, NRAO
1970	ACIDO CIANIDRICO	HCN	3,4 MILLIMETRI	10	36 PIEDI, NRAO
1970	X-OGENO	?	3,4 MILLIMETRI	8	36 PIEDI, NRAO
1970	CIANOACETILENE	HC_3N	3,3 CENTIMETRI*	4	140 PIEDI, NRAO
1970	IDROGENO	H_2	1060 ANGSTROM	2	NRL
1970	ALCOOL METILICO	CH_3OH	35,9 CENTIMETRI*	3	140 PIEDI, NRAO
1970	ACIDO FORMICO	HCOOH	18,3 CENTIMETRI	1	140 PIEDI, NRAO
1971	MONOSOLFURO DI CARBONIO	CS	2,0 MILLIMETRI*	20	36 PIEDI, NRAO
1971	FORMAMMIDE	NH_2CHO	6,5 CENTIMETRI*	2	140 PIEDI, NRAO
1971	SOLFURO CARBONILICO	OCS	2,5 MILLIMETRI*	1	36 PIEDI, NRAO
1971	OSSIDO DI SILICIO	SiO	2,3 MILLIMETRI*	2	36 PIEDI, NRAO
1971	CIANURO DI METILE	CH_3CN	2,7 MILLIMETRI	1	36 PIEDI, NRAO
1971	ACIDO ISOCIANICO	HNCO	3,4 MILLIMETRI*	1	36 PIEDI, NRAO
1971	ACIDO ISOCIANIDRICO (?)	HNC?	3,3 MILLIMETRI	3	36 PIEDI, NRAO
1971	METILACETILENE	CH_3CCH	3,5 MILLIMETRI	1	36 PIEDI, NRAO
1971	ACETALDEIDE	CH_3CHO	28,1 CENTIMETRI	1	140 PIEDI, NRAO
1972	TIOFORMALDEIDE	H_2CS	9,5 CENTIMETRI	1	210 PIEDI, CSIRO
1972	ACIDO SOLFIDRICO	H_2S	1,8 MILLIMETRI	7	36 PIEDI, NRAO
1972	METANIMINA	CH_2NH	5,7 CENTIMETRI	1	210 PIEDI, CSIRO

Le molecole interstellari individuate, per ora 26, sono qui elencate in ordine cronologico secondo la data di scoperta, assieme alla lunghezza d'onda e il telescopio usati per osservarle la prima volta. Un asterisco vicino alla lunghezza d'onda, indica che in seguito la molecola corrispondente è stata osservata anche ad altre lunghezze d'onda. Due molecole meritano particolare attenzione: il cianogeno (CN), scoperto nel 1970 alla lunghezza d'onda di 2,6 millimetri nella banda delle microonde, e l'ossido di carbonio, scoperto nel 1971 alla lunghezza d'onda di circa 1400 angstrom nell'ultravioletto. Solo queste due molecole sono state osservate in due regioni completamente diverse dello spettro. Tra le molecole riportate, otto sono state scoperte solo in una sorgente: la nuvola interstellare Sagittario B2, eccezionalmente densa e massiccia, associata a una regione H II non lontano dal centro galattico. È strano che finora l'idrogeno (H_2), la piú semplice delle molecole interstellari, sia stato osservato in due sole nuvole; infatti la forma stabile dell'idrogeno, in tutte le nuvole interstellari sufficientemente dense, dovrebbe rendere possibile la scoperta di altre molecole. L'indicazione « 100 pollici, MWO », si riferisce al telescopio riflettente di 100 pollici dell'Osservatorio di Monte Wilson; « 84 piedi, LL » al radiotelescopio di 84 piedi del Lincoln Laboratory dell'Istituto di Tecnologia del Massachusetts; « 20 piedi HCO » al radiotelescopio di 20 piedi dell'Osservatorio Hat Creek dell'Università della California; « 140 piedi, NRAO » al radiotelescopio di 140 piedi del National Radio Astronomy Observatory di Green Bank; « 36 piedi, NRAO », al radiotelescopio di 36 piedi del National Radio Astronomy Observatory situato nell'Osservatorio Nazionale di Kitt Peak; « NRL », a un razzo *Aerobee* lanciato dal Naval Research Laboratory; « 210 piedi, CSIRO », al radiotelescopio di 210 piedi della Commonwealth Scientific and Industrial Research Organization, in Australia.

Osservazioni dell'intensità delle transizioni fanno sí che sia possibile dedurre le temperature all'interno della nuvola. In questo modo, l'ossido di carbonio ha portato a temperature comprese tra i 5 e i 25 °K nelle nuvole di polvere estese, e dell'ordine di 100 °K nelle nuvole molecolari che si trovano intorno alle regioni H II.

L'ammoniaca è una sonda particolare per le condizioni fisiche delle nuvole interstellari, avendo differenti tipi di livelli energetici. La molecola è un tetraedro con l'atomo di azoto situato in un vertice sopra il piano formato dai tre atomi di idrogeno; l'atomo di azoto, può oscillare avanti e indietro rispetto al piano. Ne consegue che ogni livello di energia rotazionale, J, si divide in due livelli ravvicinati a causa dell'oscillazione. Le transizioni tra questi due livelli, sono note come transizioni di inversione e si osservano nelle microonde a 1,3 centimetri. Per l'ammoniaca, un secondo numero quantico, K, si riferisce alla componente del momento angolare attorno all'asse di simmetria della molecola. (I due numeri quantici si scrivono assieme: J,K.) La ammoniaca può cambiare di un'unità il valore di J per emissione o assorbimento di fotoni, mentre le collisioni possono far cambiare J di piú di una unità. D'altra parte K non può cambiare valore per emissione o assorbimento di fotoni. Le collisioni tra particelle in un gas, provocano transizioni solo tra livelli in cui K vale $3n + 1$ e $3n - 1$, o tra livelli in cui K vale $3n$ (con n intero). Le specie in cui K vale $3n + 1$ e $3n - 1$ si chiamano para-ammoniaca; quelle in cui K è $3n$, orto-ammoniaca.

Il numero di molecole di ammoniaca nei vari livelli di energia rotazionale può essere direttamente stimato attraverso l'osservazione dell'intensità di emissione dalle transizioni inverse. Alcuni livelli rotazionali (come il livello 2,1) decade in 20 secondi a un livello piú basso (come il livello 1,1) con la emissione spontanea di un fotone. Perché sia possibile osservare una riga spettrale dal livello 2,1 forte come quella che è stata vista, è necessario che operi una forte eccitazione di collisione, onde compensare la diseccitazione per radiazione. Perciò la densità deve essere molto alta. In realtà, nella sorgente di ammoniaca del Sagittario B2, potrebbe essere necessaria una densità fino a 10^9 particelle al centimetro cubo. Le transizioni tra altri stati (come 1,1 e 2,2) esistono solo come risultato di collisioni, che però a queste alte densità si verificano molto frequentemente. Questi stati vengono perciò portati in equilibrio termico con le particelle dell'ambiente in un periodo di tempo che è dell'ordine dell'anno. Il rapporto tra le molecole che popolano lo stato 1,1 e quelle che popolano lo stato 2,2 (cioè il rapporto di intensità che si osserva tra le righe spettrali), dovrebbe essere controllato solo dalla temperatura delle particelle che collidono. Ragionamenti di questo tipo hanno portato a temperature comprese tra 20 e 80 °K, per varie nuvole nella regione del centro galattico.

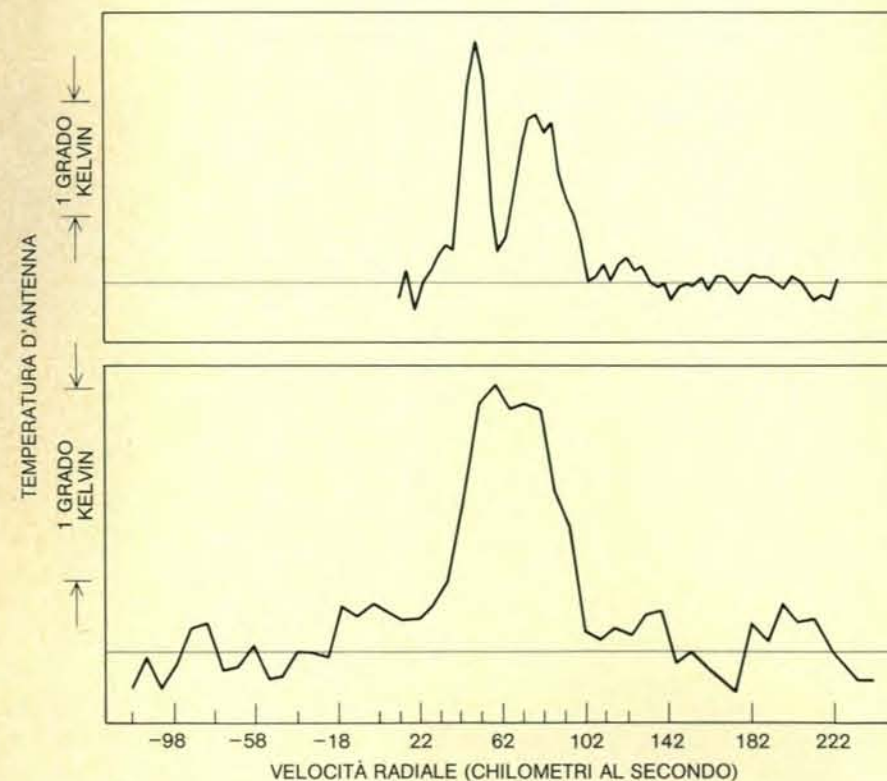
Le molecole o i loro elettroni possono cambiare lo stato di moto dando origine a varie righe dello spettro elettromagnetico. Ogni molecola tende a ruotare attorno al proprio asse di simmetria caratteristico; nel caso del solfuro di carbonio, una molecola biatomica semplice, la rotazione è continua (*a*). Cambiamenti nella rotazione fanno sí che la molecola assorba o irraggi energia a lunghezze d'onda che giacciono normalmente nella regione dello spettro delle microonde. Gli atomi della molecola vibrano uno rispetto all'altro come se fossero uniti da una molla (*b*). Anche il moto vibrazionale può cambiare, facendo sí che la molecola emetta o assorba radiazioni infrarosse. Anche gli elettroni dei vari atomi possono saltare da un livello energetico a un altro, cambiando la configurazione delle orbite elettroniche attorno al nucleo della molecola (*c*). Queste transizioni provocano emissione o assorbimento di radiazione nelle regioni ultraviolette o visibili dello spettro. In questo caso l'atomo di carbonio è a sinistra e l'atomo di zolfo a destra; la molecola si vede prima della transizione nel primo livello eccitato (*in alto*) e dopo la transizione nel livello fondamentale (*in basso*). Le linee di contorno rappresentano la probabilità di trovare un elettrone nelle varie parti della nuvola elettronica. I contorni interni sono virtualmente gli stessi per un atomo di carbonio e di zolfo isolati che per la molecola; procedendo verso i contorni esterni la probabilità di trovare un elettrone decresce.

Dall'ammoniaca si può derivare una altra temperatura che si può ottenere dalla presenza di due differenti specie di ammoniaca: l'orto-ammoniaca (alla quale appartiene lo stato 3,3) e la para-ammoniaca (alla quale appartengono gli stati 1,1 e 2,2). Le transizioni tra queste specie si possono verificare solo se si cambia lo spin del nucleo dell'idrogeno (il protone). Questo processo può verificarsi solo quando le molecole si rompono e si riformano. Il tempo medio che intercorre tra collisioni che fanno cambiare lo spin è probabilmente inferiore al milione di anni. Perciò il rapporto tra le intensità delle righe spettrali di molecole nello stato 1,1 o 2,2 rappresenta la temperatura che c'era nella nuvola circa un milione di anni fa. Si trova che queste temperature sono maggiori delle temperature attuali: questo fatto indica che le nuvole si raffreddano quando si contraggono.

Si potrebbe pensare che le osservazioni delle righe molecolari, che forniscono tante informazioni sulla temperatura e la densità delle nuvole interstellari, potrebbero anche dare una misura accurata della abbondanza delle molecole. Questo invece, non succede per una serie di ragioni. Una è che l'intensità con cui una nuvola di molecole assorbe o emette energia elettromagnetica a una certa lunghezza d'onda, non dipende solo dal numero totale di molecole che sono presenti, ma anche da quale frazione di esse si trova nei livelli energetici delle transizioni osservate. La frazione si può calcolare se sappiamo che le molecole si trovano in equilibrio termico con l'ambiente. Tuttavia la maggior parte delle molecole interstellari non si trova in equilibrio termico. Gli esempi piú spettacolari di non equilibrio sono i maser a ossidrile e i maser ad acqua. Un altro problema è che la banda ricevente di un radiotelescopio è piuttosto ampia, e quindi spesso non è in grado di risolvere piccole zone di emissione molecolare. Il risultato è che l'intensità di queste zone è sottostimata. Tuttavia, si ritiene che le quantità dedotte per la maggior parte delle 26 molecole che si conoscono, sia precisa entro un fattore 10. D'altra parte, le quantità di radicale ossidrilico e di acqua che sono stimate quando questi agiscono come un maser interstellare, possono essere affette da un errore di un fattore 100 o anche 1000. Nonostante queste incertezze, sono emerse importanti conclusioni, che permettono di capire la chimica interstellare.

Composizione chimica delle nuvole

Le nuvole interstellari in cui sono presenti quantità rivelabili di molecole, sono composte quasi interamente di particelle di polvere e di idrogeno molecolare. Le particelle di polvere si osservano direttamente perché arrossano e attenuano la luce delle stelle dello sfondo. A causa delle grandi difficoltà sperimentali, l'idrogeno molecolare è stato osservato direttamente in due sole nuvole, regioni molto sottili situate davanti alle stelle Perseo Xi e Scorpione Delta, che oscurano le stelle dello sfondo di circa una magnitudine (all'incirca 2,5 volte). Al momento non si sa se queste nuvole contengono altre molecole. Considerazioni teoriche, tuttavia, indicano che le nuvole che provocano una estinzione della luce stellare dello sfondo superiore a 1,5 magnitudini circa, contengono idrogeno sotto forma molecolare piuttosto che atomica. La velocità con cui le molecole di idrogeno si formano sulle particelle di polvere supera quella con cui esse vengono dissociate dalla radiazione ultravioletta. In generale si verifica che le molecole che si osservano nella banda ultravioletta o visibile dello spettro, non si vedono mai, nella stessa regione anche nel campo radio. La ragione di questo fatto è che la velocità alla quale una molecola emette o assorbe fotoni alle radiofrequenze, è molto minore della velocità alla quale emette o assorbe fotoni nel visibile o nell'ultravioletto. Perciò è necessario che molte

I livelli energetici di una molecola sono definiti dal numero quantico J, una misura del momento rotazionale della molecola. La molecola di solfuro di carbonio impiega in media quattro ore per decadere dal primo livello energetico al livello fondamentale. Le linee verticali tra i livelli energetici $J = 3$, $J = 2$, $J = 1$ e $J = 0$ rappresentano le transizioni del solfuro di carbonio che i radioastronomi hanno osservato nello spazio interstellare. L'unità di misura cm^{-1}, inverso del numero d'onda, è proporzionale all'energia e corrisponde al numero di lunghezza d'onda comprese in un centimetro. Le transizioni si misurano in termini dei cambiamenti di energia, e l'inverso del numero d'onda è un'indicazione diretta della quantità di energia che è coinvolta. Per esempio, una differenza di energia dell'inverso di cinque numeri d'onda corrisponde a una transizione che ha una lunghezza d'onda di due millimetri (regione delle microonde).

Gli spettri del solfuro di carbonio nella nuvola interstellare Sagittario B2, sono illustrati per due transizioni della molecola. Una data riga spettrale ha una certa ampiezza che è determinata in parte dalla turbolenza; quanto più è stretta la riga spettrale, tanto più è piccola la turbolenza vera e propria e perciò tanto più è piccola la regione della nuvola che emette la riga. La riga emessa dal solfuro di carbonio quando decade dal secondo livello energetico al primo (da $J = 2$ a $J = 1$), mostra due picchi distinti, che indicano che la riga è emessa da due regioni discrete che hanno differenti velocità radiali (*in alto*). D'altra parte la riga risultante dal decadimento dal primo livello energetico al livello fondamentale ($J = 1$ a $J = 0$) è emessa su un intervallo molto maggiore di velocità (*in basso*). Collisioni tra le particelle portano le molecole nei livelli energetici più alti; queste immagini indicano che all'interno della grande nuvola di solfuro di carbonio ci sono due regioni dense più piccole. Le linee orizzontali corrispondono al livello medio del segnale nullo che è affetto da fluttuazioni dovute al rumore.

più molecole giacciano sulla linea della visuale perché si possano produrre segnali osservabili nelle radioonde.

Ogni nuvola che contiene una quantità sufficiente di molecole per potere essere osservata alle lunghezze d'onda radio, deve contenere anche tanta polvere da far sí che le stelle del fondo, davanti alle quali le molecole si osservano come righe di assorbimento nel visibile o nell'ultravioletto, siano completamente oscurate. Il risultato è che le molecole rivelate nel visibile e nell'ultravioletto (come l'idrogeno molecolare e CH o CH^+), si osservano solo nelle nuvole a bassa densità, dove la estinzione della luce stellare è piccola e la densità delle particelle non supera le 100 particelle al centimetro cubo. D'altra parte, le molecole osservate alle lunghezze d'onda radio si vedono solo nelle nuvole più scure e più dense, dove la densità delle particelle è compresa tra le 100 e le 10^9 particelle al centimetro cubo, e l'estinzione supera probabilmente le 50 magnitudini (un fattore superiore a 10^{20}).

In queste dense nuvole di idrogeno molecolare e di polvere, si osservano solamente tracce di altre molecole. Il motivo è che le molecole contengono sí altri atomi, oltre all'idrogeno, ma questi, in scala universale, sono molto meno abbondanti dell'idrogeno. Ogni 10 000 atomi di idrogeno, ce ne sono tre o quattro di ossigeno, due di carbonio, uno o due di azoto e uno di zolfo. Qui, tuttavia, emerge un fatto importante. Se la scala delle abbondanze universali (determinata con precisione dagli spettri atomici che si osservano in stelle vicino al Sole) vale anche per le nuvole interstellari in generale, allora le molecole osservate, da sole, costituiscono quasi tutti gli atomi di carbonio che devono essere presenti. Esse devono costituire anche una frazione considerevole (forse il 30 %) di tutti gli atomi d'ossigeno. Invece, forse non più che lo 0,0001 % degli atomi d'azoto presenti compaiono nelle molecole osservate. Quasi tutti gli atomi di azoto si trovano probabilmente sotto forma di molecola biatomica (N_2), che non è stata osservata e che costituisce la maggior parte dell'atmosfera terrestre.

Due importanti conclusioni seguono da questi fatti. La prima è che la formazione di molecole da atomi che sono presenti, deve essere un meccanismo molto efficiente nelle nuvole interstellari. La seconda è che la chimica interstellare sembra favorire la produzione di molecole organiche, cioè di molecole contenenti carbonio. Nello spazio interstellare sono state trovate

tutte le molecole organiche più semplici, mentre, nonostante siano state fatte ricerche accurate non sono state rivelate molte delle molecole inorganiche anche più semplici. Finora non si sa spiegare perché la chimica interstellare è prevalentemente organica.

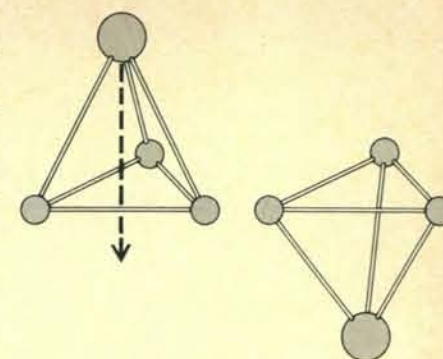
Molecole e radiazione di fondo

Nel 1965, Arno A. Penzias e R. W. Wilson dei laboratori Bell, scoprirono che l'universo è immerso in un campo di radiazione uniforme, che ha una temperatura equivalente ai 3 °K nel campo delle microonde. Se si potesse dimostrare che questa radiazione ha uno spettro di « corpo nero », cioè che ha una temperatura equivalente a 3 °K a tutte le lunghezze d'onda, si potrebbe logicamente spiegare questo fatto assumendo che essa è quanto rimane dell'esplosione primordiale che segnò l'inizio della attuale fase di espansione dell'universo. D'altra parte, se si trovassero forti deviazioni dallo spettro di corpo nero, questa origine dovrebbe essere accantonata. Dal 1969, le osservazioni delle molecole interstellari hanno avuto un ruolo importante per la soluzione di questo importante problema. Fin dal 1969 si sa che la formaldeide interstellare, come l'ossidrile e l'acqua, è eccitata con procedimenti molto strani. P. Palmer dell'Università di Chicago, B. M. Zuckerman dell'Università del Maryland, L. E. Snyder dell'Università della Virginia e David Buhl del National Radio Astronomy Observatory, osservarono una transizione della formaldeide alla lunghezza d'onda di 6 centimetri. Essi videro la transizione come riga di assorbimento in alcune nuvole di polvere, su una regione del cielo che non conteneva sorgenti conosciute di radiazione con uno spettro continuo. La radiazione assorbita doveva perciò essere la radiazione cosmica di fondo a 3 °K. Poiché l'osservazione della formaldeide non presenta molte delle difficoltà tecniche che sono associate alla misura diretta del fondo, il fatto costituisce una delle migliori conferme dirette dell'esistenza della radiazione cosmica di fondo.

Misure di eccitazione di CN, CH e CH^+, hanno recentemente permesso di fare delle stime dell'intensità della radiazione di fondo a lunghezze d'onda radio inferiori ai 3 millimetri. In questa zona dello spettro non esistono altre informazioni attendibili. Nelle nuvole rarefatte in cui si osservano queste tre molecole, la frequenza di collisione con altre particelle è cosí bassa, che le loro popolazioni relative dei pochi livelli energetici più bassi sono largamente governate dalla frequenza con cui interagiscono con i fotoni della radiazione cosmica di fondo. Misure delle popolazioni relative, dedotte dalle osservazioni di due o più transizioni che coinvolgono i rispettivi livelli energetici, portano alla intensità della radiazione cosmica di fondo alle lunghezze d'onda di queste transizioni. Il CN dà che l'intensità è proprio attorno ai tre °K, alla lunghezza d'onda di 2,63 millimetri. Il CH e il CH^+ conducono a risultati analoghi, anche se meno definitivi, alle lunghezze d'onda di 0,56 e 0,36 millimetri.

Abbondanze isotopiche

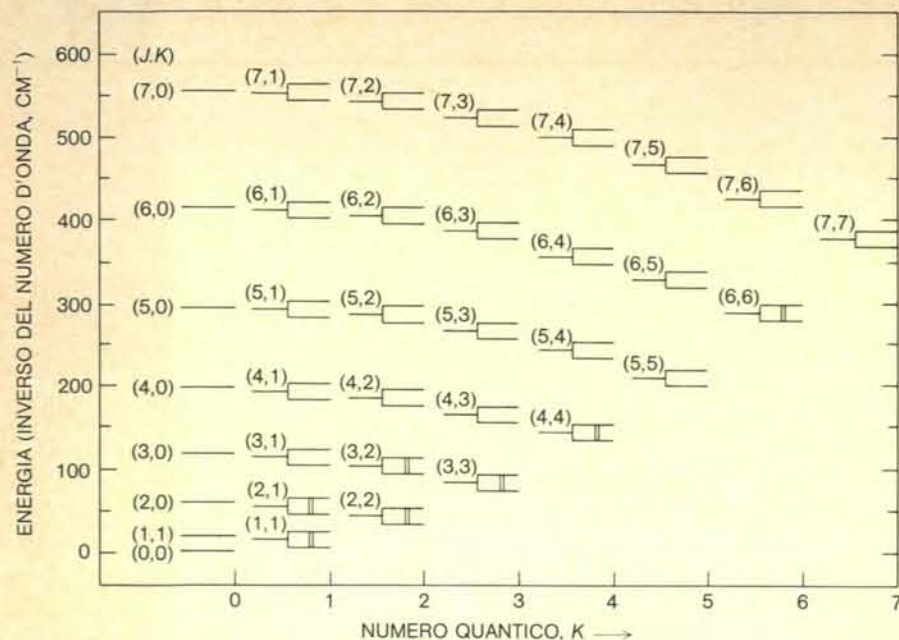
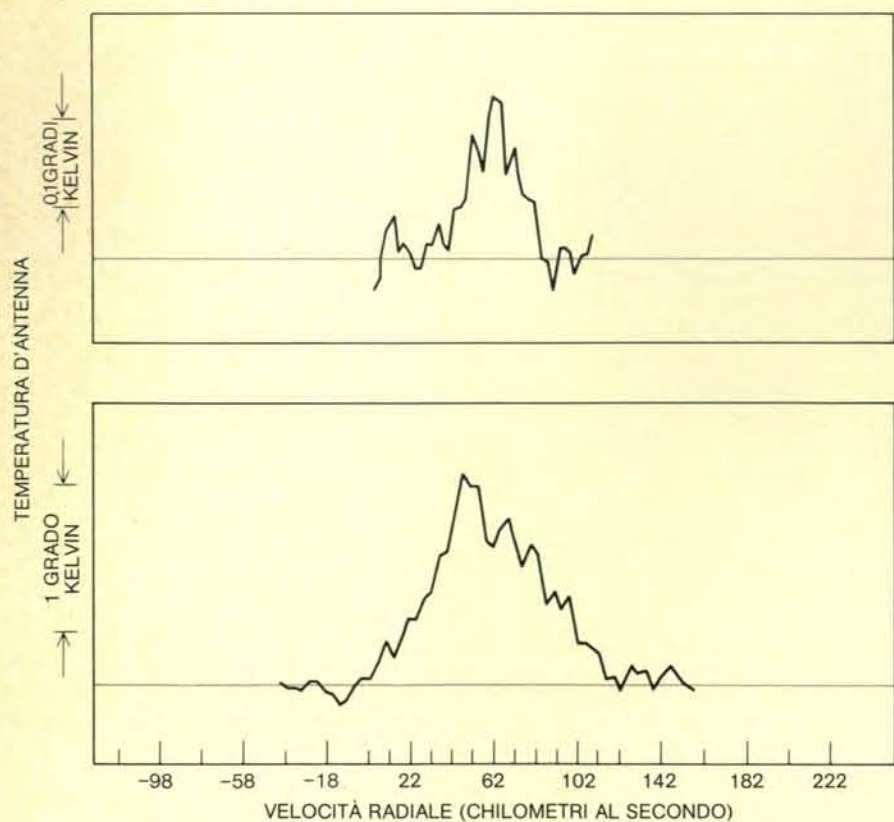
Sulla Terra e nel Sole, nelle comete e nei meteoriti, si trovano in proporzioni ben definite i differenti isotopi degli elementi chimici. Come ho già detto, il rapporto di abbondanza tra il carbonio-12 e il carbonio-13 è di 89 : 1. Il rapporto tra ossigeno-16 e ossigeno-18 è di 490: 1, tra azoto-14 e azoto-15 è di 270 : 1 e tra zolfo-32 e zolfo-34 è di 22 : 1. Nei tipi di stelle che bruciano la loro energia nucleare in processi che coinvolgono carbonio, azoto e ossigeno, i rapporti isotopici di questi elementi sono alterati. Il rapporto tra carbonio-12 e carbonio-13 può essere addirittura di 4 : 1, come abbiamo visto, e il rapporto tra azoto-14 e azoto-15 può diventare molto grande. Nelle stelle che bruciano elio, il carbonio viene utilizzato in modo tale da consumare tutto il carbonio-13.

Le stelle riciclano continuamente la loro materia attraverso il mezzo interstellare. Si creano dal gas interstellare, sottopongono il gas a processi di combustione nucleare e quindi restituiscono il gas allo spazio interstellare emettendo materia o esplodendo in supernove. Perciò i rapporti isotopici nel sistema solare dovrebbero essere accidentali: il risultato di processi di non equilibrio o il mescolamento di gas provenienti da altre sorgenti. Il sorprendente risultato che ci ha fornito lo studio di molecole interstellari contenenti differenti isotopi, è che i rapporti isotopici interstellari sembrano essere uguali a quelli sulla Terra. L'unica eccezione sembra essere la regione del centro galattico, in cui il rapporto tra carbonio-12 e carbonio-13 è apparentemente il doppio che in tutti gli altri posti. Questo fatto potrebbe suggerire che, in proporzione, più materia interstellare è passata nel centro galattico attraverso stelle che bruciano carbonio, o perché in questa regione c'è meno gas collegato a stelle, o perché, delle stelle presenti, un numero più grande è del tipo

La molecola di ammoniaca (NH_3) è un tetraedro con un atomo di azoto nel vertice al di sopra del piano formato da tre atomi di idrogeno. L'atomo di azoto può oscillare da una parte (*sinistra*) all'altra (*destra*) del piano. Come risultato, ciascun livello di energia rotazionale, J, è diviso in due livelli ravvicinati. Le transizioni tra i due livelli sono transizioni di inversione e si hanno alla lunghezza d'onda di 1,3 cm.

massiccio a combustione di carbonio. Il fatto che la maggior parte del mezzo interstellare assomiglia alla Terra nei rapporti isotopici, sembrerebbe indicare che la chimica interstellare è cambiata poco nei cinque miliardi di anni che sono passati da quando è nato il nostro pianeta. Ci sono due spiegazioni possibili. Forse non esistono regioni (tranne il centro galattico) in cui sono molto più numerose le stelle a combustione di carbonio che le stelle a combustione d'elio. Oppure il gas interstellare deve essersi ben mescolato su grandi distanze e in un tempo paragonabile alla breve vita delle stelle più calde.

Cicli vitali delle molecole

Come abbiamo visto, fu una sorpresa per gli astronomi scoprire molecole semplici quali CN, CH e CH^+. Essi ritenevano che l'abbondanza molecolare dovesse essere estremamente esigua, perché è altamente improbabile che due atomi si uniscano per dar luogo a una molecola, e perché una qualsiasi molecola sarebbe stata rapidamente distrutta dalla inesorabile radiazione ultravioletta interstellare. D'altra parte, è ovvio che bisogna elaborare nuove teorie per spiegare l'esistenza di molecole organiche complesse, dato che queste vengono distrutte ancora più velocemente, e che di esse si osservano quantità molto maggiori che di CN, CH e CH^+.

Negli ultimi anni sono stati proposti vari meccanismi per spiegare la formazione di molecole. Essi si possono dividere in cinque categorie. La prima prevede che le molecole si formino da atomi che collidono e che rimangono

I livelli energetici dell'ammoniaca differiscono da quelli del solfuro di carbonio a causa dell'oscillazione dell'atomo di azoto. Il secondo numero quantico, K, si riferisce alla componente del momento attorno all'asse di simmetria della molecola. I due numeri quantici si scrivono assieme: J, K. L'ammoniaca può cambiare di un'unità il valore di J emettendo o assorbendo un fotone; le collisioni possono far cambiare J per più di un'unità. Il secondo numero quantico K può invece cambiare soltanto per collisione.

Gli spettri dell'ammoniaca nel Sagittario B2 danno informazioni sulla densità all'interno della sorgente. Il numero di molecole di ammoniaca che si trova nei vari livelli di energia rotazionale si può stimare osservando le transizioni di inversione. Alcuni livelli rotazionali, come il livello 2,1 (*in alto*), decade in un tempo di soli 20 secondi a un livello più basso, per esempio al livello 1,1 (*in basso*), con l'emissione spontanea di un fotone. Perché sia possibile osservare una intensa riga spettrale dal livello 2,1 come quella che è stata scoperta, è necessario che vi sia un'alta frequenza di collisione che contrasti il decadimento spontaneo. Perciò la densità deve essere alta. Nel Sagittario B2 potrebbe essere necessaria una densità perfino di 10^9 particelle al centimetro cubo.

attaccati uno all'altro, mentre si muovono nel mezzo gassoso interstellare. La seconda prevede che le molecole si formino sulla superficie di particelle di polvere, da atomi o molecole che vengono in contatto su queste superfici. La terza, che le molecole nascano dalla evaporazione o dalla decomposizione di particelle di polvere, quando queste sono colpite da particelle energetiche o da fotoni, o incontrano onde d'urto o altri fenomeni che provocano riscaldamento. La quarta prevede che le molecole si generino dalle collisioni atomiche nella densa atmosfera delle stelle e che quindi siano espulse nello spazio interstellare. La quinta, che le molecole si formino nell'ambiente denso che circonda «una nebulosa presolare», cioè durante la fase del collasso di una protostella.

Le molecole devono nascere alla stessa velocità con la quale muoiono, altrimenti la vita di una nuvola molecolare sarebbe limitata e in realtà potrebbe essere molto corta. I meccanismi principali che distruggono le molecole interstellari sono la fotodissociazione da parte della radiazione ultravioletta e il congelamento sulla superficie di particelle di polvere. Recentemente, L. J. Stief e i suoi colleghi del Goddard Space Flight Center, hanno studiato la velocità con cui le molecole sono fotodissociate utilizzando i risultati di una elegante esperienza di laboratorio e misure dell'intensità della radiazione ultravioletta eseguite con satelliti artificiali. L'importante risultato è che molecole come acqua, ammoniaca, formaldeide e solfuro carbonilico si dissociano in meno di 100 anni se non sono schermate dalla polvere e sono esposte al flusso medio di radiazione interstellare ultravioletta. Nella scala astronomica del tempo si tratta di un intervallo di tempo veramente breve.

Tuttavia, le nuvole di polvere sono uno schermo molto efficace alla radiazione ultravioletta. Una tipica nuvola di polvere con un diametro di un anno luce e una densità di 1000 molecole di idrogeno al centimetro cubo, è in grado di attenuare la luce di circa quattro magnitudini e la radiazione ultravioletta di un fattore che forse raggiunge 10^{24}. La conseguenza è che la vita media di una molecola può essere aumentata fino a 10 milioni di anni. Dobbiamo perciò concludere che o le molecole interstellari si sono formate all'interno delle nuvole di polvere, dove esse si osservano oggi, o che esse sono state trasportate all'interno di queste nuvole con qualche protezione, forse sulla superficie di particelle di polvere.

La densità delle particelle di polvere all'interno delle grandi nuvole di polvere è tale che una molecola si condenserà dal gas sulla superficie di una di queste particelle in circa 100 000 anni. In queste nuvole le temperature sono troppo basse perché sia probabile che le molecole evaporino di nuovo nello stato gassoso. Un periodo di 100 000 anni è notevolmente piú corto della vita media di queste nuvole. Perciò, indipendentemente dal modo in cui le molecole si sono formate, esse devono in qualche modo rigenerarsi piú volte dalle particelle di polvere durante la vita della nuvola. È possibile che questa rigenerazione sia opera dei pochi raggi cosmici o fotoni di radiazione ultravioletta che riescono a penetrare nella nuvola, o da sorgenti invisibili di radiazione infrarossa, tipo protostelle, che si trovano nella nuvola.

La situazione è molto piú complicata nelle dense nuvole molecolari associate a regioni H II. A causa della forte densità della nuvola, le molecole si possono attaccare alle particelle di polvere in soli 100 anni. Tuttavia, le maggiori temperature che esistono all'interno di queste nuvole, fanno sí, probabilmente, che una parte significativa delle molecole evapori di nuovo nello stato gassoso. Mentre le molecole si trovano sulle particelle di polvere, possono intervenire reazioni di superficie che generano molecole nuove e piú complicate. Questi effetti potrebbero spiegare perché nelle regioni H II si trova una varietà maggiore di molecole complesse che nelle nuvole di polvere relativamente rarefatte.

Per adesso possiamo solo affermare che non c'è alcun accordo tra gli astronomi molecolari in favore di qualcuno dei cinque meccanismi proposti per la formazione delle molecole; forse nessuno di essi va del tutto bene. C'è però un insieme di modi in cui questi meccanismi potrebbero agire per formare molecole. Oggi, alcune scelte tra le varie possibilità sembrano emergere dalle osservazioni; pare infatti che la sorgente diretta della maggior parte delle molecole interstellari, non possa essere l'atmosfera delle stelle. Le molecole complesse, nel loro passaggio verso il mezzo interstellare, non potrebbero sopravvivere al passaggio attraverso la calda corona dell'atmosfera stellare, anche se si trovasse un meccanismo opportuno di espulsione. Per di piú si sa dalla teoria (e in pochi casi dall'osservazione) che i tipi di stelle che hanno atmosfere sufficientemente fredde per formare molecole interstellari hanno un rapporto tra carbonio-12 e carbonio-13 compreso tra 4 : 1 e 10 : 1, in contrasto con il rapporto interstellare di 89 : 1.

Vi sono alcune evidenze, anche se non troppo conclusive, che le molecole non vengono solo dalla disintegrazione di particelle di polvere interstellare. Se cosí fosse, anche le particelle di polvere non si sarebbero formate nello spazio interstellare, ma avrebbero dovuto essere espulse dalle stelle. In linea di principio ciò è certamente possibile, ma è in disaccordo con il fatto che solo stelle con un basso rapporto tra carbonio-12 e carbonio-13 possono produrre efficacemente particelle di polvere. Le molecole interstellari che provengono dalla disintegrazione di particelle di

ATOMO	RAPPORTO ISOTOPICO TERRESTRE	RAPPORTO ISOTOPICO INTERSTELLARE	MOLECOLA INTERSTELLARE E REGIONE
CARBONIO-12/CARBONIO-13	89:1	CIRCA 50:1 SE OSSIGENO-16/OSSIGENO-18=488:1	H_2CO, CENTRO GALATTICO
		CIRCA 89:1 SE OSSIGENO-16/OSSIGENO-18=870:1	H_2CO, CENTRO GALATTICO
		CIRCA (89±15):1	H_2CO, AL
		CIRCA (82±15):1	CH^+, ZETA OPHIUCHI
OSSIGENO-16/OSSIGENO-18	488:1	488:1 SE CARBONIO-12/CARBONIO-13=50:1	H_2CO, CENTRO GALATTICO
		488:1 SE CARBONIO-12/CARBONIO-13=50:1	CO, CENTRO GALATTICO
		870:1 SE CARBONIO-12/CARBONIO-13=89:1	H_2CO, CENTRO GALATTICO
		(390±100):1	OH, CENTRO GALATTICO
OSSIGENO-16/OSSIGENO-17	2700:1	ALMENO 2700:1	OH, CENTRO GALATTICO
AZOTO-14/AZOTO-15	270:1	MAGGIORE DI 70:1	NH_3, CENTRO GALATTICO
		(230±70):1	HCN, ORIONE
ZOLFO-32/ZOLFO-34	22,5:1	(24±5):1	CS, CENTRO GALATTICO E ALTROVE
ZOLFO-32/ZOLFO-33	125:1	MAGGIORE DI 100:1	CS, ORIONE

I rapporti isotopici dei vari atomi nello spazio interstellare a confronto con gli stessi rapporti isotopici sulla Terra. Quando si usano molecole diverse dall'ossido di carbonio e dalla formaldeide, il paragone si può fare in maniera relativamente semplice per ossigeno, azoto e i due isotopi dello zolfo. Nel centro galattico, tuttavia, la grande abbondanza di ossido di carbonio e di formaldeide confonde l'interpretazione. Le righe spettrali si «saturano», cioè l'abbondanza non è piú proporzionale alla intensità della riga spettrale. In questo caso, si può ottenere solo il rapporto di carbonio e ossigeno assieme, cioè il rapporto tra carbonio-13 e ossigeno-18 e carbonio-12 e ossigeno-16. Si trova che questo rapporto è 1,8 volte il rapporto terrestre. Non si conosce alcun processo nucleare nelle stelle che modifichi il rapporto tra ossigeno-16 e ossigeno-18; per di piú, questo rapporto, se misurato attraverso il radicale ossidrilico, sembra avere lo stesso valore terrestre. Perciò si ritiene che, nella regione del centro galattico, il valore non terrestre del rapporto di carbonio e ossigeno sia dovuto al valore 50 (non terrestre) del rapporto tra carbonio-12 e carbonio-13. Questo rapporto potrebbe essere dovuto alla combustione nucleare del carbonio nelle stelle giganti, che possono essere piú numerose nel centro galattico. Nelle altre parti della Galassia, anche il rapporto del carbonio sembra essere lo stesso che sulla Terra. Poiché tutti i rapporti isotopici interstellari sembrano essere molto vicini ai valori terrestri, la chimica del mezzo interstellare deve essere rimasta costante per un tempo almeno pari all'età della Terra.

polvere dovrebbero condividere un basso rapporto tra carbonio-12 e carbonio-13, cosa contraria alle osservazioni. Inoltre, le molecole non si osservano in prevalenza in regioni in cui non c'è evidenza di onde d'urto, particelle energetiche subatomiche o altri fattori che potrebbero decomporre le particelle di polvere per riscaldamento.

Perciò sembra che il primo, o il secondo, o il quinto meccanismo, o una loro combinazione, siano responsabili della formazione delle molecole interstellari. Processi diversi possono certamente verificarsi in tipi diversi di nuvole molecolari. In linea di principio, in ogni tipo di nuvola possono avere luogo reazioni sulla superficie delle particelle di polvere, a una velocità che dipende dalla densità, temperatura e composizione chimica delle particelle. Anche altri meccanismi primari possono essere i maggiori responsabili della formazione di molecole sulle particelle di polvere. Per esempio, George H. Herbig, dell'Osservatorio Lick, ha suggerito che le molecole si possono creare in nebulose presolari che si trovano nelle nuvole dense che circondano le regioni H II. Si sa che le regioni H II sono regioni di formazione stellare: molte nebulose presolari dovrebbero esservi presenti. Queste nebulose possono espellere abbastanza efficacemente molecole e polvere nelle nuvole molecolari circostanti. Le molecole e la polvere sono espulse dalla pressione di radiazione infrarossa proveniente dalla protostella centrale o da un eccesso di momento rotazionale, che si sviluppa durante gli ultimi stadi del collasso della nebulosa presolare. Nelle protostelle le reazioni nucleari non sono ancora iniziate; in questi oggetti, i rapporti isotopici dovrebbero essere gli stessi di quelli che si osservano nello spazio interstellare.

Solo il primo e il secondo meccanismo sembrano funzionare per spiegare la formazione di molecole nelle nuvole di polvere meno dense, che sono lontane dalle regioni H II. Anche a questi meccanismi, tuttavia, sono connessi vari problemi. Se si considerano solo le reazioni superficiali, rimane la domanda: come si sono formate le particelle di polvere? In teoria, esse potrebbero essere state create altrove e essere venute a riposo in seguito al passaggio attraverso nuvole più dense del mezzo interstellare circostante. Oppure le particelle di polvere potrebbero essersi formate in lunghi periodi di tempo attraverso processi di accrescimento, iniziati con coppie di atomi unitisi a formare molecole biatomiche. I dettagli teorici di questo processo sono piuttosto complicati. Per quanto le molecole biatomiche si possano formare a un ritmo adeguato mediante questo processo, sembra che le successive collisioni con atomi del gas circostante provocherebbero la distruzione delle molecole già formate, piuttosto che crearne di nuove con più atomi. Solamente quando le molecole sono formate da almeno una dozzina di atomi, le collisioni nel gas sarebbero in grado di creare molecole più grandi.

Qualunque sia il meccanismo responsabile della formazione di molecole nelle nuvole di polvere, sembra certo che, date le basse densità e le basse temperature che ci sono in queste nuvole, la velocità con cui le molecole interagiscono con le particelle di polvere deve essere molto più bassa che nelle nuvole più dense, vicino alle regioni H II. Inoltre, nelle nuvole di polvere, probabilmente non esistono sorgenti primarie efficaci di molecole come le nebulose presolari. In queste nuvole a bassa densità, sembra che le molecole più complesse (come il cianoacetilene e l'alcool metilico) siano presenti in quantità minori delle molecole più semplici (come l'ossidrile, il solfuro di carbonio e la formaldeide), di quanto non succeda nelle nuvole più dense vicino alle regioni H II. Se questo sospetto fosse confermato da successive esperienze, tale differenza in composizione chimica potrebbe indicare che, in un tipo di sorgente, il meccanismo di produzione molecolare è differente da quello che si verifica nell'altro tipo.

Ulteriori problemi

I meccanismi che ho descritto finora sembrano in grado di formare molecole interstellari, nelle quantità osservate, in regioni adeguatamente schermate dalla radiazione ultravioletta. Tuttavia non spiegano né le abbondanze relative delle molecole osservate, né l'assenza di certe altre molecole. Quello che si sa sulla velocità con cui le molecole sono dissociate dalla radiazione ultravioletta, indica che non c'è alcun rapporto tra queste velocità e le quantità che si osservano. La conclusione inevitabile è che funzionano meccanismi di formazione che sono molto particolari e molto poco capiti. Quasi certamente coinvolgono reazioni sulla superficie di particelle di polvere. Queste reazioni sono probabilmente importanti anche nelle nuvole vicine alle regioni H II dove le nebulose presolari possono essere la causa principale della formazione molecolare; le molecole, dopo essere state proiettate dalla protostella nella nuvola circostante, devono continuare a interagire rapidamente con le particelle di polvere. Sfortunatamente si sa poco delle reazioni di superficie, anche nel caso che le superfici siano ben definite. Non si conosce, al momento, la composizione delle particelle di polvere interstellare: essa può variare da regione a regione secondo le condizioni locali. Col passare del tempo sono stati proposti come candidati alla composizione delle particelle di polvere, grafite, grafite ricoperta da acqua gelata, ammoniaca e metano, e un miscuglio di grafite, silicato e ferro. È evidente che le particolari molecole che si formano per catalisi su queste superfici, devono dipendere fortemente dalla composizione delle particelle di polvere.

Ho concluso intenzionalmente questa discussione con un tono un po' pessimistico. Appare infatti chiaro che occorrono molte discipline scientifiche per risolvere i problemi della chimica interstellare. Si può dire che si è già sviluppata una nuova disciplina: l'astrochimica. La nostra conoscenza delle molecole interstellari che, come si è visto è piena di lacune, ha potuto fare dei passi lunghissimi, servendosi del lavoro comparato in campi tradizionalmente non astronomici. Gli spettroscopisti delle microonde, hanno contribuito a dare informazioni sulle esatte frequenze alle quali bisogna ricercare nuove molecole. I fotochimici hanno misurato la probabilità che i fotoni ultravioletti distruggano particolari molecole. Per molti anni, gli esperimenti sulla formazione di semplici molecole organiche, sono stati portati avanti da chi era interessato a scoprire come nasce la vita. Con la scoperta di almeno quattro molecole interstellari, che si ritengono indispensabili per la formazione di molecole biologiche (acqua, formaldeide, acido cianidrico e cianoacetilene), si possono portare elementi in favore o contro l'affascinante problema che si chiede se si può scoprire nello spazio la vita primordiale, o almeno gli amminoacidi che sono i suoi fondamentali costituenti.

Nonostante questi problemi, che manterranno il campo dell'astrochimica in uno stato di particolare eccitazione per un certo tempo, la spettroscopia molecolare si è rivelata il maggiore strumento astronomico. I livelli energetici delle molecole sono di solito molto più ravvicinati dei livelli energetici degli atomi singoli: perciò le molecole si adattano particolarmente allo studio delle regioni fredde che si trovano tra le stelle. Questo fatto, sommato alla maggiore efficacia che le lunghezze d'onda delle righe molecolari hanno di penetrare regioni di polvere e di gas, ci ha permesso di esplorare le nuvole di polvere e gli stadi primari di formazione di stelle entro esse.

Osservando i geni in azione

Il microscopio elettronico permette di osservare i singoli geni durante la loro trascrizione nel rispettivo RNA e nel momento in cui questo RNA è tradotto in proteina

di O. L. Miller jr.

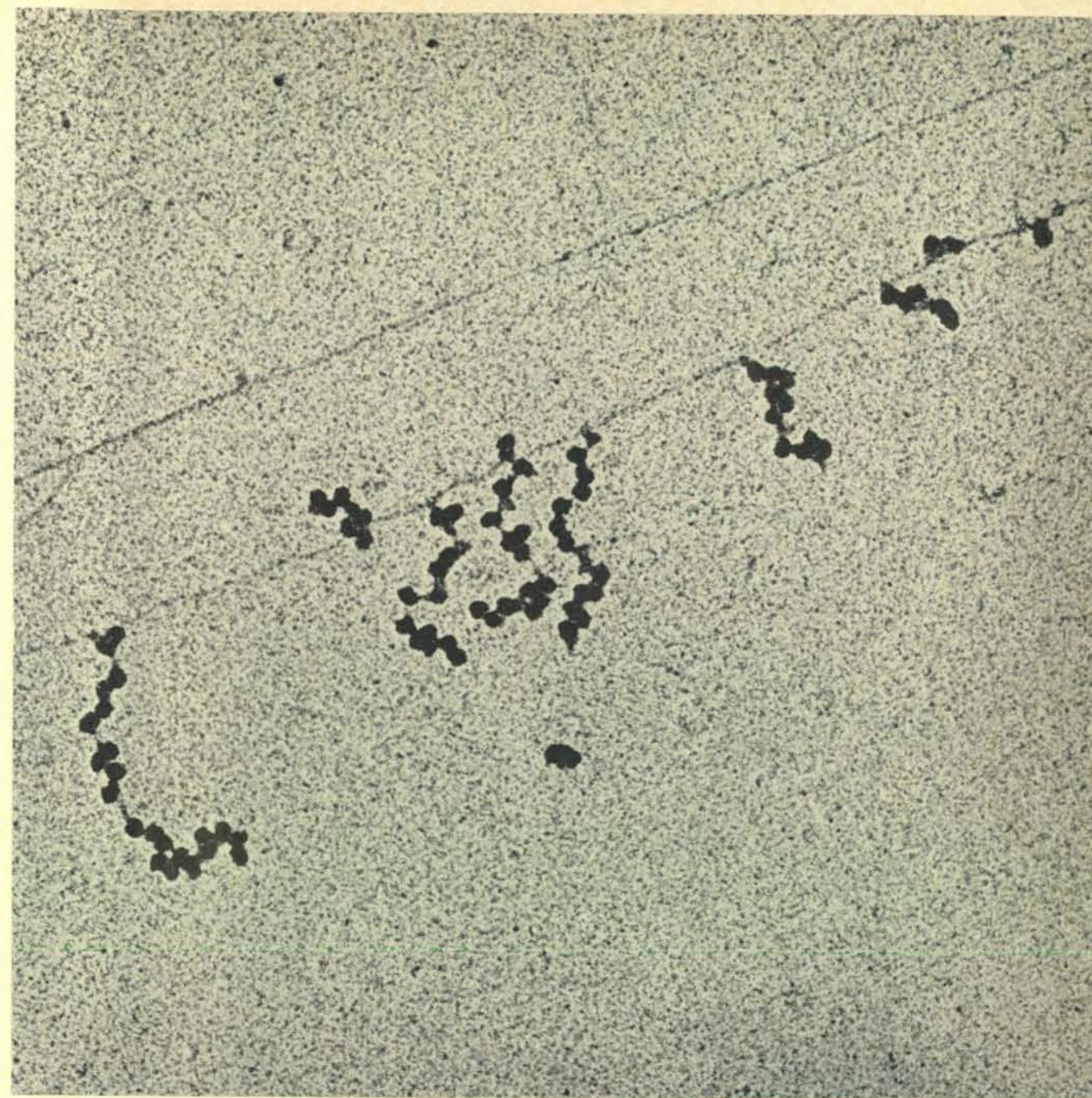
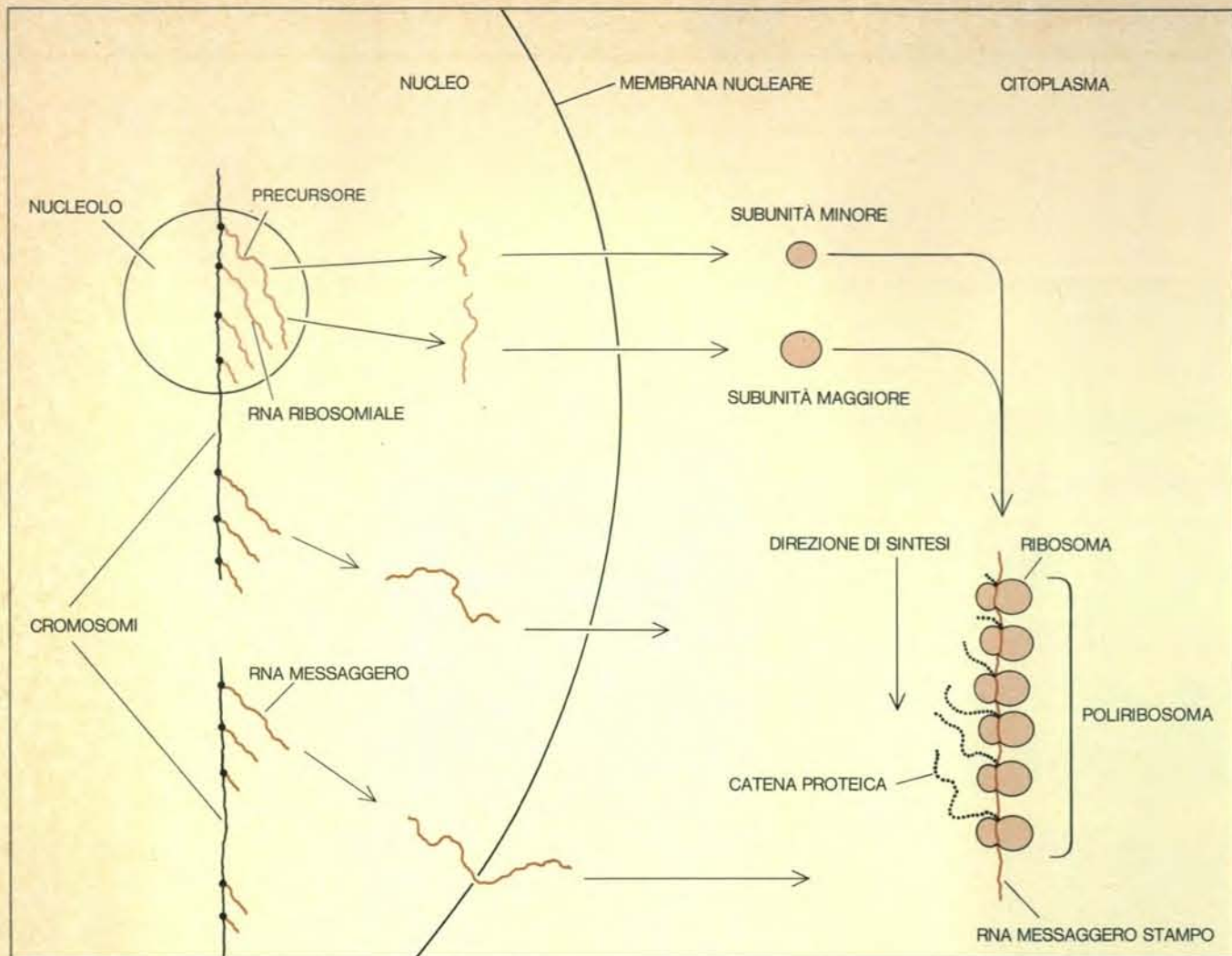
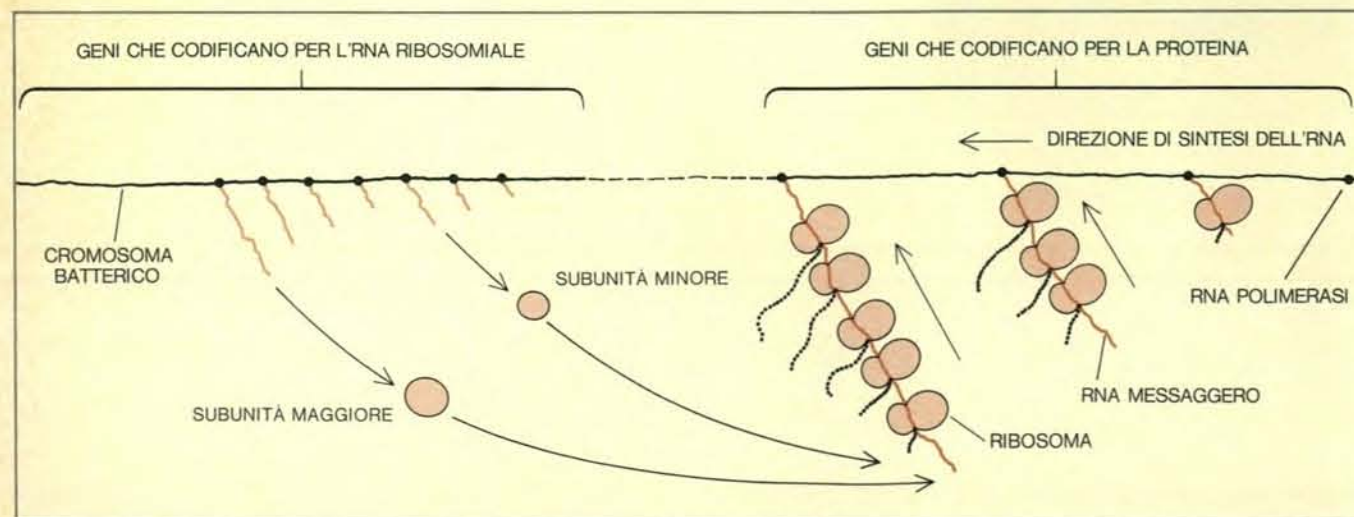
A partire dalla metà degli anni '50, un importante obiettivo della biologia è stato quello di documentarsi e di arricchire di particolari le conoscenze che già si possedevano sul cosiddetto dogma centrale della genetica: cioè che il DNA è il materiale ereditario, che l'informazione in esso contenuta è codificata nelle sequenze delle subunità che costituiscono i geni e che quest'informazione viene trascritta nell'RNA e quindi tradotta in proteina. Grazie a una varietà di tecniche biochimiche e genetiche notevoli furono identificate molte fasi del processo di trasferimento dell'informazione assieme alle sostanze e agli elementi cellulari interessati in esse, fu elaborato il « dizionario » DNA-proteina e determinato l'effetto, sulla trascrizione e sulla traduzione, di differenti condizioni e agenti estranei. Tuttavia il quadro sempre meglio precisato che si andava costruendo era di necessità basato, in gran parte, su prove indirette e di varia provenienza: vi erano infatti scarse testimonianze dirette del funzionamento dei singoli geni.

Nel 1967, presso l'Oak Ridge National Laboratory, i miei colleghi e io tentammo per la prima volta di fotografare al microscopio elettronico singoli geni in azione. Il successo che ottenemmo indica che la microscopia elettronica è potenzialmente un utile strumento per lo studio della citogenetica a livello molecolare. La prima soddisfazione che ci diede il nostro lavoro fu l'osservazione che molte delle nostre immagini mostrano una straordinaria rassomiglianza con gli schemi della trascrizione e della traduzione che, per anni, sono stati pubblicati su riviste tecniche o scientifiche. In altre parole queste immagini tendono a confermare ciò che era stato proposto in via teorica dopo aver considerato i dati quantitativi che erano stati raccolti con notevole fatica.

Il trasferimento dell'informazione dall'acido desossiribonucleico (DNA) genico alla proteina si effettua attraverso una serie di tappe intermedie che comportano la trascrizione del DNA in tre tipi di acido ribonucleico (RNA). Uno di questi RNA è l'RNA messaggero che funge da stampo su cui le subunità amminoacidiche si riuniscono in proteina nella cosiddetta fase di traduzione. Un altro è l'RNA ribosomiale, un costituente di quelle strutture che sono note come ribosomi e sulle quali ha luogo la traduzione. Il terzo è l'RNA di trasporto, che trasporta appunto le subunità amminoacidiche della proteina sullo stampo costituito dall'RNA messaggero, nel punto più opportuno.

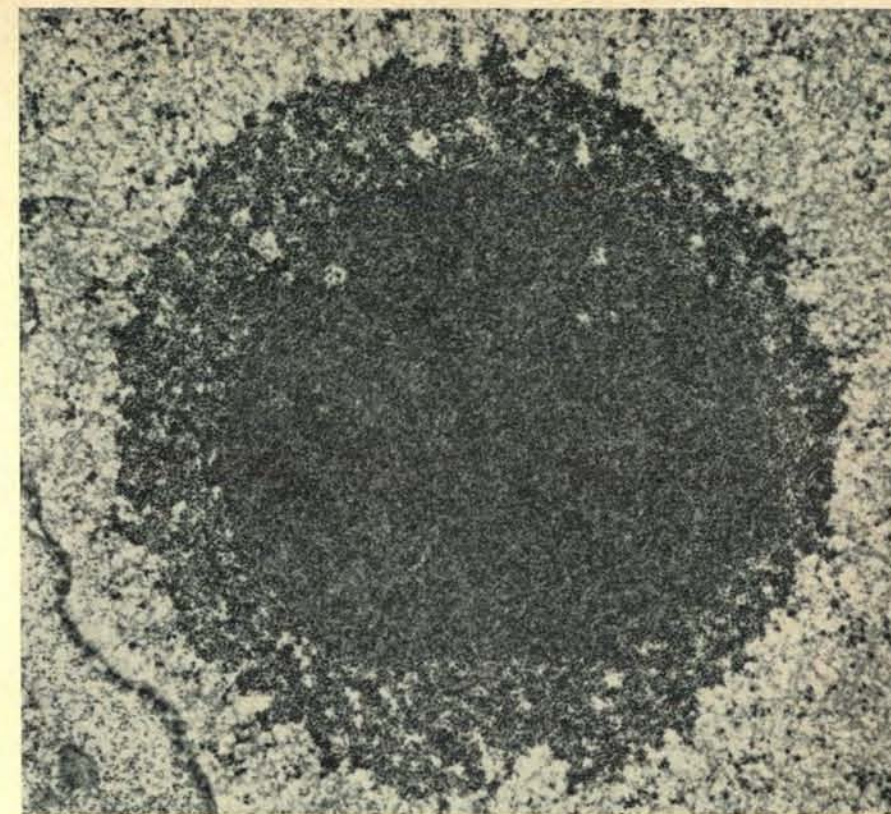
Le fasi che compongono la trascrizione del DNA in RNA e la traduzione dell'RNA nella proteina sono essenzialmente uguali in tutte le cellule viventi, ma i loro rapporti temporali e spaziali nelle cellule degli eucarioti, il cui DNA è confinato per la massima parte del ciclo cellulare all'interno di un nucleo delimitato da una membrana, sono diversi da quelli delle cellule dei procarioti, che non sono dotate di questo tipo di nucleo. Nei procarioti, che includono le numerose specie di batteri, la trascrizione e la traduzione dell'RNA messaggero hanno luogo contemporaneamente e nella stessa sede. Nelle cellule degli eucarioti, che comprendono tutti gli organismi dotati di nucleo dagli unicellulari, come il paramecio, alle cellule dei vertebrati superiori, i due processi sono separati come tempo e come spazio: la trascrizione del DNA nei vari RNA, che intervengono poi nella sintesi proteica, avviene nel nucleo, quindi gli RNA migrano attraverso la membrana nucleare nel citoplasma, che costituisce la massa più vistosa della cellula e che contiene il meccanismo per la sintesi proteica (*si vedano le figure a pagina 76*).

Decidemmo di cominciare a studiare il meccanismo genetico di una cellula eucariotica: l'oocita, o elemento germinale femminile, di anfibi quali le rane o le salamandre. Questo tipo di oocita si presta allo studio dell'attività dei geni per un certo numero di ragioni, la più importante delle quali, per noi, era che ci avrebbe dato una buona probabilità di vedere moltissimi geni facilmente identificabili: quelli che codificano per le due grosse molecole di RNA che si trovano nei ribosomi. Nella massima parte delle altre cellule eucariotiche questi geni sono presenti in ammassi localizzati nei nucleoli, corpi densi situati in punti specifici di certi cromosomi. Negli oociti degli anfibi viene prodotto un gran numero di copie dei geni per l'RNA ribosomiale e queste copie si trovano in centinaia di nucleoli liberi nel nucleo e indipendenti dai cromosomi. Il risultato di questa « amplificazione genica » per l'RNA ribosomiale è che un unico nucleo dell'oocita equivale, per quanto riguarda un locus genetico ben identificato, a molte centinaia di nuclei cellulari tipici e che i geni amplificati non sono mescolati e confusi con altri geni sui cromosomi.

Si traggono due altri vantaggi dallo studio degli oociti di anfibio. Uno è che queste cellule, mentre si accrescono, sintetizzano attivamente l'RNA messaggero e i loro cromosomi si trovano nello stadio « a spazzola », dilatati e srotolati, in modo da poter vedere i particolari strutturali del DNA quando sta per essere trascritto nell'RNA messaggero. Il secondo vantaggio è che questi oociti hanno una dimensione notevole: in alcune specie raggiungono,

Microfotografia elettronica di gene batterico in azione ingrandito 150 000 volte, eseguita, come le altre dello stesso articolo, dall'autore e dai suoi collaboratori. L'immagine elettronica va interpretata secondo lo schema che appare a destra. Si notano due segmenti di cromosoma del batterio *Escherichia coli*, di cui l'inferiore è attivo, cioè il suo DNA viene trascritto nell'RNA messaggero e quest'ultimo viene tradotto in proteina. Nella microfotografia si possono vedere molecole di RNA polimerasi, l'enzima che catalizza la trascrizione del DNA nell'RNA: una, all'estrema destra, è in corrispondenza del punto in cui approssimativamente ha inizio la trascrizione di tutte le molecole di RNA. I filamenti di RNA messaggero successivamente trascritti (*in colore nel disegno*) si staccano in effetti verso sinistra; il più lungo, a sinistra nella microfotografia, è stato il primo a essere sintetizzato. A mano a mano che il filamento di RNA si allunga, i ribosomi si avvicinano e si muovono lungo di esso, verso il cromosoma, traducendo l'RNA in proteina (*non rappresentata*).

I processi di trascrizione e di traduzione si svolgono rispettivamente nel nucleo e nel citoplasma delle cellule eucariotiche. L'RNA ribosomiale (*in colore chiaro*), trascritto dai geni presenti nel nucleolo, si forma come precursore che viene scisso all'interno del nucleolo stesso per formare le due grosse molecole di RNA che entrano a far parte delle due unità del ribosoma (in realtà è improbabile che gli RNA provenienti dallo stesso precursore vadano a finire nello stesso ribosoma). Gli RNA messaggeri (*in colore scuro*), trascritti da altri geni, servono da stampi su cui i poliribosomi riuniscono le subunità amminoacidiche in catene proteiche. I due tipi di RNA, man mano che vengono sintetizzati, formano complessi con le proteine. Negli oociti di anfibio, un numero elevato di copie di geni per l'RNA ribosomiale è presente anche nei nucleoli extracromosomici.

Trascrizioni e traduzioni si svolgono simultaneamente nelle cellule procariotiche che non possiedono un vero nucleo. L'RNA ribosomiale (*in colore chiaro*) viene trascritto a partire da due geni, apparentemente contigui su certi segmenti dell'unico cromosoma batterico. L'RNA messaggero (*in colore scuro*) viene trascritto da geni posti altrove e subito tradotto in proteine dai ribosomi.

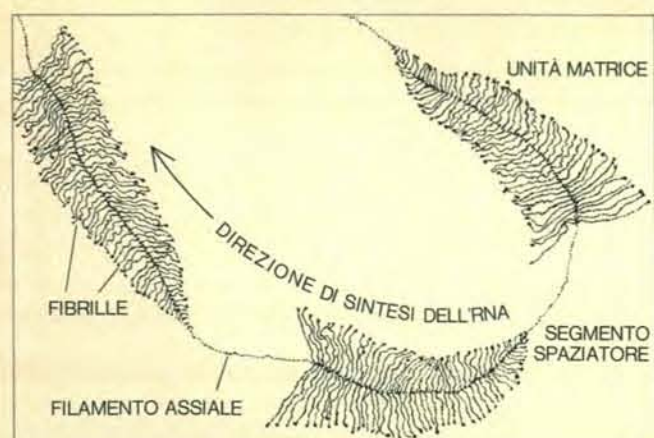
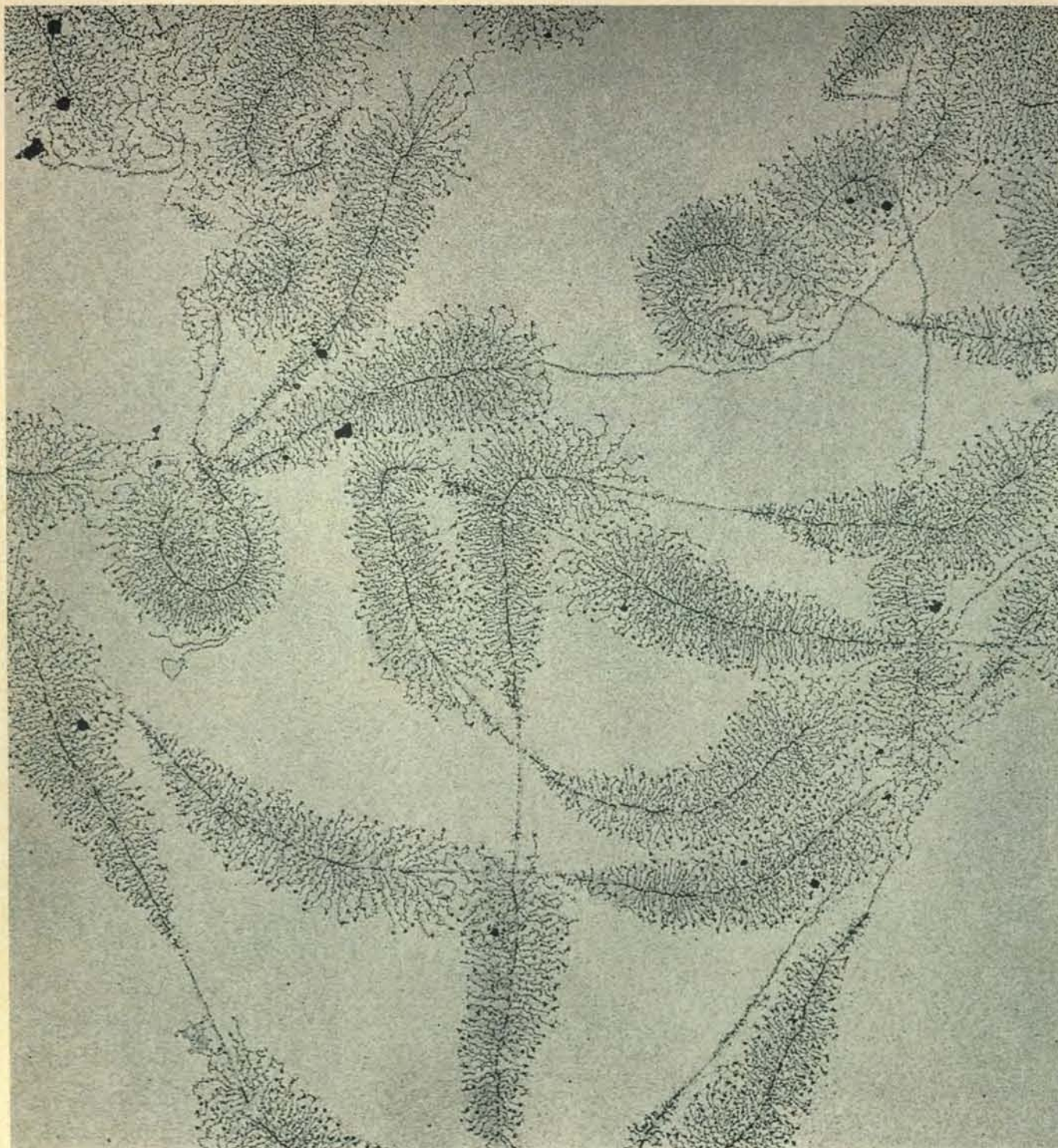
quando sono maturi, 2 mm di diametro e possiedono nuclei del diametro di anche 1 mm. Con un microscopio a basso potere risolutivo e con una pinzetta finissima si possono facilmente isolare i singoli nuclei e manipolarne il contenuto per ottenere un preparato da esaminare al microscopio elettronico.

Barbara Beatty e io abbiamo notato che, isolando il contenuto del nucleo di un oocita e mettendolo subito in acqua distillata, gli strati esterni granulari dei nucleoli extracromosomici si disperdevano rapidamente, per cui la porzione centrale, compatta e fibrosa, che speravamo contenesse i geni che stavamo cercando, si poteva espandere. Abbiamo poi sospeso in una soluzione di formalina la porzione centrale detta cuore di questi nucleoli, per fissarli, e li abbiamo centrifugati in apposite provette, in modo che andassero a disporsi su una sottile pellicola di carbonio, stesa su un retino metallico per microscopio elettronico. Bisognava che il centrifugato si distendesse bene, senza deformarsi e senza che andasse distrutto il quadro dell'attività genica che stavamo ricercando. Siamo riusciti in questo intento rendendo idrofile le pellicole di carbonio, cioè in grado di attrarre acqua, e lavando il retino, prima di asciugarlo, in acqua contenente un agente che riducesse la tensione superficiale. Abbiamo quindi reso visibile il campione con un metallo pesante, per dargli un piú forte contrasto sul cammino del fascio di elettroni, e abbiamo eseguito alcune micrografie elettroniche delle porzioni centrali dei nucleoli, dopo che esse si erano srotolate.

Ognuna di queste porzioni appariva come una lunga fibra a forma di anello ingarbugliato e deformato, lungo il quale si trovavano delle strutture che si ripetevano e che abbiamo chiamato « unità matrici » (*si veda la figura nella pagina seguente*). Ognuna di queste unità ha una lunghezza di circa 2,4 μ; tra una e l'altra vi è un segmento di filamento nudo che, generalmente, è un poco piú corto. Ogni matrice consiste di una serie di fibrille che si estendono ai lati del filamento lungo; la lunghezza di queste fibrille aumenta regolarmente lungo il filamento da un'estremità della matrice all'altra, formando una struttura a forma di freccia. Tutte le matrici di un filamento centrale del nucleolo hanno la stessa polarità, cioè tutte le « frecce » puntano nella stessa direzione. Trattando i campioni con coloranti specifici e con enzimi, abbiamo potuto identificare il lungo filamento assiale come un complesso di DNA e proteina e le fibrille

Nucleolo extracromosomico di oocita del tritone *Triturus viridescens*, in sezione sottile, ingrandito 15 000 volte. Si nota una parte centrale filamentosa e una corticale granulare. È visibile, in basso a sinistra, una parte della membrana nucleare con il citoplasma all'esterno. Il cuore del nucleolo contiene numerose copie di geni per l'RNA ribosomiale.

laterali come complessi di RNA e proteina. Mediante autoradiografia con costituenti dell'RNA marcati con isotopi radioattivi, abbiamo trovato che le matrici stavano sintetizzando attivamente RNA. I segmenti spaziatori tra le varie unità non venivano apparentemente trascritti. Le molecole nascenti di RNA, nelle fibrille, formavano complessi con la proteina prodotta nel citoplasma e quindi migrata nel nucleo, dove si associava alle molecole di RNA a mano a mano che venivano sintetizzate.

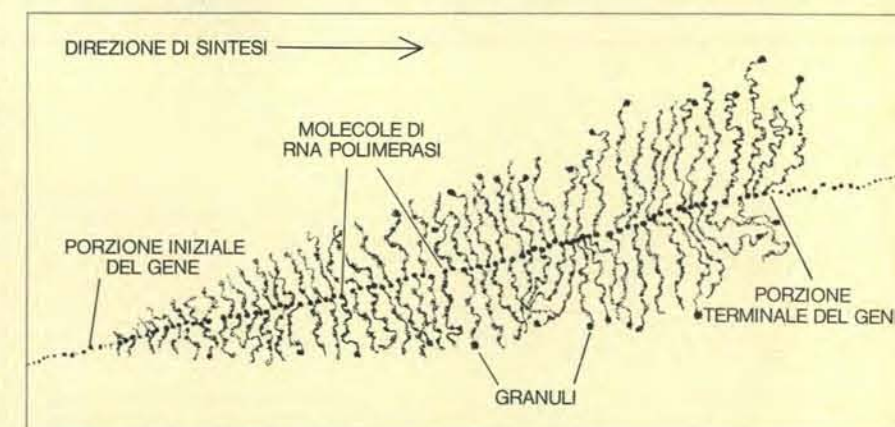
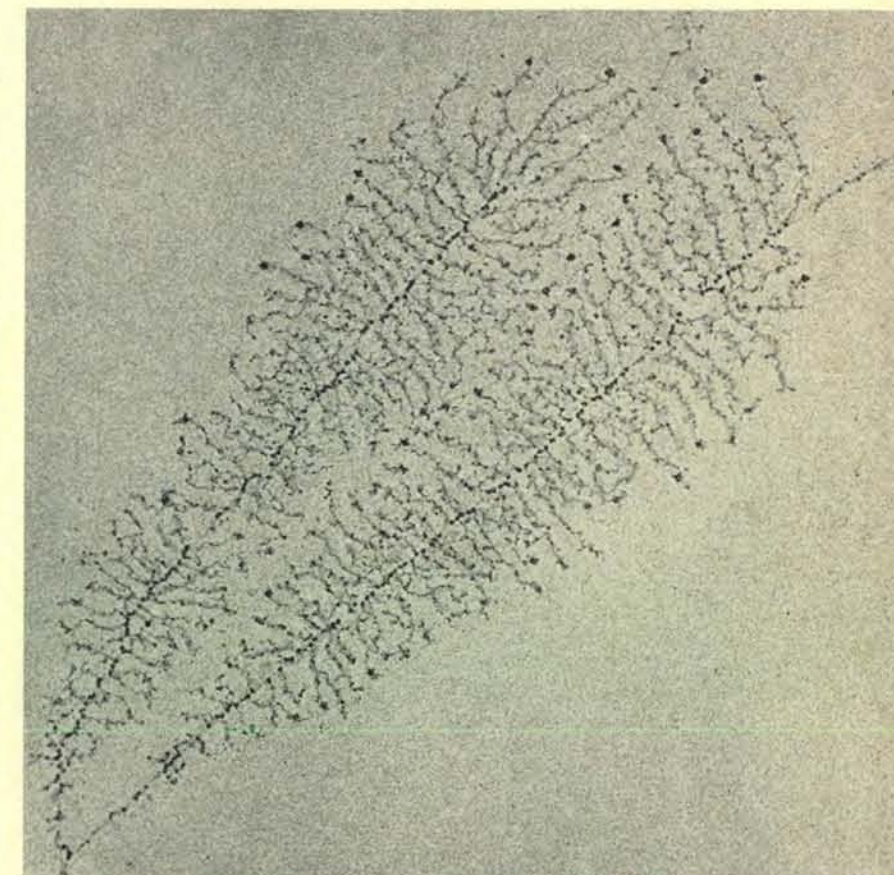
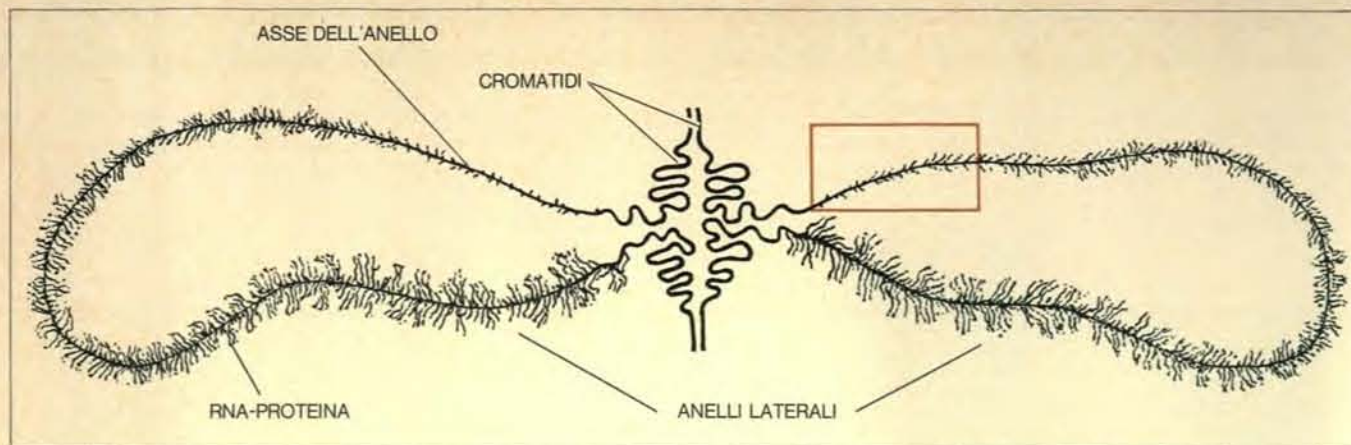
Proprio mentre stavamo compiendo queste osservazioni, ci giunse notizia di alcuni dati biochimici molto pertinenti. Numerosi ricercatori riferirono che il DNA nei nucleoli extracromosomici degli oociti di anfibio conteneva sequenze di subunità che erano complementari all'RNA dei ribosomi presenti nel citoplasma, confermando cosí che il DNA nucleolare era la fonte d'informazione dell'RNA ribosomiale. Le varie tappe del trasferimento di quest'informazione furono precisate in molti laboratori. Ne è risultato che sia nei mammiferi, sia negli anfibi, le due grosse molecole di RNA che si trovano nei ribosomi non vengono trascritte separatamente a partire dal DNA, ma vengono ricavate da un unico precursore (studi piú recenti hanno dimostrato che questo è vero, probabilmente, per tutti gli eucarioti). Questo precursore ha un peso molecolare di circa $2,7 \cdot 10^6$ negli anfibi; quindi, per codificare una simile molecola, sarebbe necessaria una doppia elica di DNA di circa 2,7 μ. Le matrici che abbiamo osservate avevano all'incirca una lunghezza di 2,4 μ, valore ragionevolmente vicino ai 2,7 μ del DNA. Poiché si è dimostrato che i nucleoli extracromosomici contengono DNA che codifica per l'RNA ribosomiale e le nostre osservazioni hanno localizzato la sintesi dell'RNA nelle unità matrici che costituiscono il « cuore » dei nucleoli, abbiamo concluso che il filamento di DNA presente in ogni unità matrice è in effetti un gene che codifica per un precursore dell'RNA ribosomiale e che le fibrille laterali contengono queste molecole in fase di sintesi.

Su alcuni filamenti che sono stati leggermente distesi durante la preparazione abbiamo potuto osservare le singole molecole dell'RNA-polimerasi (enzima che catalizza la lettura del DNA) in attività in corrispondenza di ogni sito di trascrizione (*si veda la figura a pagina 79*). Il nostro successo si deve al fatto che, quando tali geni sono attivi, dalle 80 alle 100 molecole di polimerasi trascrivono simultaneamente

Geni attivi per l'RNA ribosomiale sono situati in questa porzione centrale srotolata di un nucleolo, colorata con acido tungstico e ingrandita 26 000 volte; nello schema a sinistra sono indicati gli elementi principali. Si vede un lungo filamento assiale in forma di cerchio deformato, con una serie di unità matrici a forma di frecce per tutta la sua lunghezza. Ogni unità si compone di una serie di sottili fibrille laterali di varia lunghezza. Mediante colorazione e saggi enzimatici, il filamento assiale è stato identificato come un complesso di DNA e proteina, mentre le fibrille sono complessi di RNA e proteina. Il segmento di DNA che si trova in ogni matrice è un gene per l'RNA ribosomiale: ogni fibrilla è un precursore dell'RNA ribosomiale nell'atto di essere trascritto dal gene e unito in un complesso con una proteina. I tratti di filamento tra le unità matrici sono segmenti spaziatori inattivi, la cui funzione è ignota.

ogni gene, formando fibrille di precursori dell'RNA ribosomiale (piú la proteina), in una serie graduata di lunghezze che corrispondono a stadi successivi di sintesi. La lieve discrepanza tra la lunghezza osservata di 2,4 μ per le unità matrici e la lunghezza di 2,7 μ prevista per i geni può spiegarsi probabilmente con il fatto che il DNA si srotola per la trascrizione in corrispondenza di ciascuno dei molti punti di trascrizione; ogni rigonfiamento al di fuori della doppia elica di DNA tenderebbe ad accorciare la lunghezza totale del gene.

Eseguendo la stessa preparazione usata per i nucleoli extracromosomici, sui cromosomi di oocita di anfibio, abbiamo trovato che i minimi particolari della sintesi dell'RNA si possono osservare anche sui lunghi anelli laterali caratteristici dello stadio a spazzola di questi cromosomi (*si vedano le figure a pagina 80*). Vediamo di nuovo le molecole di RNA polimerasi disposte a stretto contatto lungo le porzioni attive del filamento, ognuna alla base di una delle fibrille nascenti di RNA-proteina, che formano un gradiente continuo lungo l'anello. Tuttavia le fibrille diventano molto piú lunghe di quelle situate sui geni nucleolari e, pertanto, devono contenere molecole di RNA molto grosse. Non conosciamo la successiva funzione delle fibrille prodotte dai cromosomi dell'oocita; presumibilmente il loro RNA dà origine a RNA messaggeri per la sintesi di proteine specifiche, ma finora non siamo in grado di assegnare nessun messaggero specifico a un qualsiasi anello.

Con la mia collaborazione, Aimée H. Bakken ha esteso le nostre ricerche sugli eucarioti alla cellula dei mammiferi, precisamente al ceppo tumorale umano HeLa, che è stato ottenuto in coltura da una ventina d'anni e che ha permesso di studiare intensamente la sintesi dell'RNA e delle proteine mediante tecniche chimiche. Il fatto che queste cellule siano molto piú piccole degli oociti (un diametro di circa 30 μ o meno), ha richiesto metodi diversi di isolamento e di dispersione del contenuto nucleare. In un tentativo di evitare noiose micromanipolazioni, abbiamo cercato di elaborare un metodo chimico per rompere le cellule e i nuclei cosí da poter spremere fuori nucleoli e cromosomi senza degradare le strutture che volevamo osservare. Dopo molte prove con vari agenti, abbiamo avuto successo trattando le cellule con basse concentrazioni di un normale detersivo per stoviglie.

Abbiamo isolato dalle cellule HeLa (presumibilmente dai nucleoli) segmenti attivi di cromosomi che abbiamo identificato provvisoriamente come geni per l'RNA ribosomiale (*si veda l'illustrazione in alto a pagina 81*). Quest'identificazione è basata innanzitutto sulla loro straordinaria somiglianza strutturale con le unità matrici degli oociti, per l'RNA ribosomiale, la cui identità poggia su solide basi. In ambedue i casi i segmenti attivi del DNA sono caricati al massimo di molecole di RNA polimerasi: perlomeno 150 molecole di queste polimerasi per gene nelle cellule HeLa. Piccoli, densi granuli compaiono alle estremità libere delle fibrille che stanno crescendo e i segmenti attivi vicini sono separati da segmenti spaziatori inattivi, che nelle cellule HeLa hanno all'incirca la stessa lunghezza dei probabili geni per l'RNA ribosomiale. I segmenti attivi sono piú lunghi nelle cellule HeLa di quanto non lo siano negli oociti, ma piú corti di quanto ci si aspetterebbe sulla base del peso molecolare del precursore del-

Molecole di RNA polimerasi, attive a livello di ogni sito di trascrizione, visibili su due geni nucleolari, distesi a seguito del processo di preparazione e ingranditi 48 000 volte. Uno di questi geni viene riprodotto nello schema. Le molecole di polimerasi appaiono come densi granuli alla base di ogni fibrilla; ognuna si muove da sinistra a destra man mano che va trascrivendo il gene. La microfotografia mostra anche granuli, a funzione finora ignota, che si vedono alla sommità delle fibrille che si stanno accrescendo.

I cromosomi a spazzola sono una forma speciale di cromosomi, visibile nei nuclei degli oociti durante il loro lungo periodo di crescita. In questo stadio le porzioni attive delle coppie di cromatidi, che costituiscono l'asse del cromosoma, sono distese a formare anelli laterali; il filamento assiale di DNA di ogni anello viene trascritto in RNA, il quale forma un complesso con una proteina. La crescente lunghezza delle molecole di RNA spiega il differente spessore dell'anello. Il rettangolo in colore mostra pressappoco il segmento di anello laterale che è visibile, ingrandito, nella micrografia in basso alla pagina.

l'RNA ribosomiale nei mammiferi: tali segmenti sono lunghi all'incirca 3,5 μ, invece dei previsti 4,5 μ. Di nuovo la discrepanza è dovuta, probabilmente, a qualche srotolamento della molecola di DNA in corrispondenza di ciascuno dei numerosi siti di trascrizione, situati a breve spazio l'uno dall'altro.

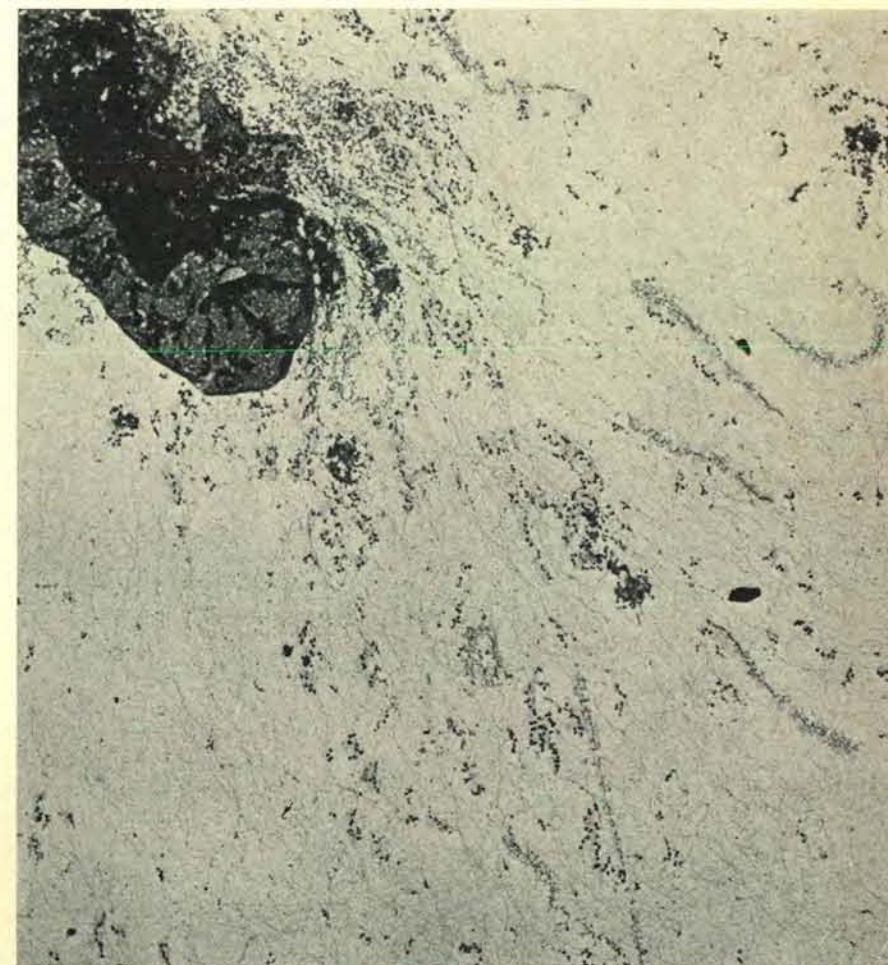
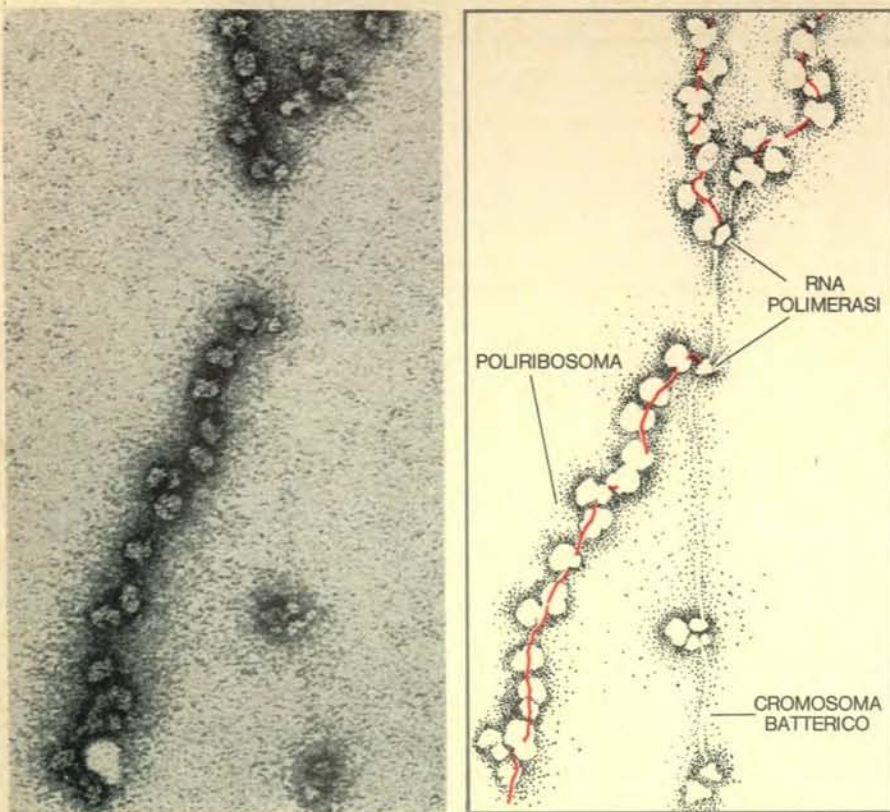
Nel considerare la sintesi dell'RNA non ribosomiale nelle cellule HeLa, sapevamo – da ricerche effettuate in altri laboratori – che piú o meno il 95 per cento dell'RNA sintetizzato in una di queste cellule, in un determinato momento, non sopravvive fino a raggiungere il citoplasma, ma viene degradato nel nucleo. Questo RNA viene chiamato RNA nucleare eterogeneo e le sue molecole hanno dimensioni variabili da circa un decimo di quella del precursore dell'RNA ribosomiale nelle cellule HeLa, a molte volte questa dimensione. Si hanno valide prove per sostenere che piccole porzioni delle molecole di RNA nucleare eterogeneo non vengono degradate e sono i precursori degli RNA messaggeri trovati nel citoplasma delle cellule HeLa.

Oltre alle ben distinte fibrille del precursore dell'RNA ribosomiale, vediamo fibrille di RNA-proteina di dimensioni varie, situate qua e là lungo cromosomi HeLa dispersi; nessuna di queste fibrille è disposta però secondo una serie graduata di lunghezze. Ciò indica che la sintesi dell'RNA si innesca molto meno frequentemente e che il livello dell'attività polimerasica è, di conseguenza, molto piú basso nei siti attivi dei cromosomi HeLa, rispetto a quanto avviene sugli anelli laterali attivi dei cromosomi a spazzola degli anfibi. Poiché l'RNA nucleare eterogeneo costituisce il 95 per cento o piú dell'RNA che viene sintetizzato nel nucleo delle cellule HeLa, è probabile che la maggior parte delle fibrille disperse di RNA-proteina che vediamo attaccate ai cromosomi HeLa contengano RNA nascente di quel tipo. Tuttavia, come nel caso dei cromosomi a spazzola, nessuna funzione genetica specifica può finora essere attribuita a una qualsiasi delle sintesi di RNA che sono state osservate.

Dopo il nostro primo lavoro con gli oociti di anfibio, ho collaborato con Charles A. Thomas jr. e Barbara A. Hamkalo, presso la Harvard Medical School, per adattare le tecniche preparatorie a una cellula procariotica: il batterio *Escherichia coli*. In seguito la Hamkalo ha continuato con me il lavoro a Oak Ridge. Disponendo di un ampio bagaglio di dati biochimici e genetici sulla trascrizione e la traduzione di vari geni di *E. coli* e di altri batteri, potevamo prevedere come sarebbero apparsi i geni batterici attivi se fossimo riusciti a fotografarli. Tuttavia il primo problema da risolvere fu quello di rendere fragile la dura parete esterna della cellula batterica, cosí da poterla rompere. Abbiamo trovato che, raffreddando rapidamente dei batteri *E. coli* in fase di crescita e facendo seguire quest'operazione da un blando trattamento con lisozima, si poteva eliminare la parete ottenendo protoplasti osmoticamente attivi, cioè cellule nude che si potevano far scoppiare per pressione osmotica immergendole in acqua. Abbiamo quindi centrifugato queste cellule come avevamo fatto con i « cuori » dei nucleoli degli oociti e abbiamo fotografato al microscopio elettronico i contenuti fuoriusciti della membrana cellulare.

Sapevamo che nei batteri, diversamente da quanto accade negli eucarioti, l'RNA messaggero nascente normalmente forma un complesso con i ribosomi in attività. Non appena un filamento di RNA messaggero è abbastanza lungo, un ribosoma si attacca a esso e comincia la sua traduzione in proteina. A mano a mano che la sintesi dell'RNA messaggero procede, con formazione di un filamento sempre piú lungo, piú ribosomi si attaccano a quest'ultimo, formando cosí una specie di collanina, chiamata poliribosoma, la quale effettua la traduzione dell'RNA a mano a mano che si allunga. Tutti i geni attivi che codificano per la sintesi proteica nei nostri campioni do-

Anello laterale di cromosoma a spazzola di un oocita di tritone, ingrandito 22 000 volte. L'estremità sottile, vicino al punto di congiungimento con l'asse del cromosoma, si trova a sinistra. La sintesi dell'RNA comincia qui, come dimostra la lunghezza crescente delle fibrille da sinistra verso destra. Alla base di ogni fibrilla sono visibili numerose molecole di RNA polimerasi.

Le unità matrici, isolate dal nucleo di una cellula HeLa, appartenente a un ceppo tumorale umano, vengono identificate come geni per l'RNA ribosomiale sulla base delle somiglianze con i geni nucleolari degli oociti di anfibio che presentano: serie graduata di fibrille leggermente spaziate con una polimerasi alla base di ognuna, densi granuli alla sommità delle fibrille e presenza di segmenti spaziatori (*non visibili*) tra i geni attivi.

Cellula di *Escherichia coli* lisata per pressione osmotica. La micrografia mostra i residui della parete cellulare (*in alto a sinistra*) e il contenuto cellulare fuoriuscito, consistente in sottili filamenti con segmenti fibrillari e collanine di granuli attaccate a essi. I filamenti sono porzioni del cromosoma batterico, le collanine di granuli, poliribosomi e i segmenti fibrillari, geni ribosomiali. L'ingrandimento è di 22 000 diametri; la microfotografia di pagina 75 rappresenta un ulteriore ingrandimento di questo materiale.

La colorazione negativa (con acetato di uranile) sottolinea il contorno del materiale estratto da una cellula di *E. coli*, mostrando che le molecole di RNA polimerasi (*schema a destra*) sono piú piccole e piú irregolari di quelle degli eucarioti. La colorazione rivela filamenti di quello che, presumibilmente, è RNA messaggero, ma non definisce bene il cromosoma. L'ingrandimento della microfotografia è di 150 000 diametri.

vrebbe dunque presentare una serie graduata di collanine di poliribosomi.

Microfotografie elettroniche del contenuto, fuoriuscito dalle cellule, di *E. coli*, mostrano masse di filamenti sottili, ad alcuni dei quali sono attaccate collanine di granuli (*si veda la figura in basso a pagina 81*). Se si tratta un campione con l'enzima desossiribonucleasi, i filamenti vengono distrutti; trattandone un altro con la ribonucleasi, le collanine di granuli scompaiono dai filamenti. Si conclude cosí che i filamenti sono porzioni del cromosoma batterico e che i granuli sono ribosomi nell'atto di tradurre RNA messaggeri. Il fatto che non vi siano poliribosomi attaccati lungo larghi tratti dei cromosomi concorda con le affermazioni che una buona parte del cromosoma di *E. coli* codifica o per specie rare di RNA oppure non viene mai trascritto. Inoltre, i batteri da noi coltivati si trovano in condizioni ottimali per cui molte tappe biosintetiche possono non avere alcuna necessità di svolgersi e i geni che codificano per gli enzimi che devono intervenire in esse possono essere repressi.

A piú forte ingrandimento appare la configurazione ben dettagliata delle regioni ricoperte dai ribosomi: i geni che codificano per la sintesi proteica (*si veda la figura a pagina 75*). Il gene appare come un sottile filamento assiale, lungo il quale vi è una serie graduata di collanine di ribosomi. In corrispondenza della giunzione di ogni collanina di ribosomi con il gene, si riesce a risolvere la molecola attiva dell'RNA polimerasi, che sta catalizzando la trascrizione. Con la colorazione negativa si può notare che gli enzimi hanno una molecola piú piccola e di forma piú irregolare di quelle delle cellule eucariotiche (*si veda la figura in questa pagina in alto*). Sfortunatamente le foto non mettono in evidenza le catene proteiche, presumibilmente associate con ogni ribosoma, nella fase di allungamento: gli amminoacidi che costituiscono queste catene sono semplicemente troppo piccoli per poter essere visibili con i nostri metodi.

Per quanto riguarda i geni di *E. coli*, che codificano per l'RNA ribosomiale, si potrebbe prevedere per essi un aspetto molto diverso da quello dei geni per la sintesi proteica. Mentre gli RNA messaggeri di geni che codificano per le proteine sono contenuti entro i poliribosomi, l'RNA ribosomiale nascente dovrebbe formare un complesso con qualche proteina, ma non con ribosomi maturi. Studi quantitativi sulla sintesi dell'RNA ribosomiale in *E. coli* in fase di rapida crescita, assieme a valutazioni in base alle quali, all'interno di ogni cromosoma di *E. coli*, non esistono piú di sette segmenti che codificano per la sintesi dell'RNA ribosomiale, hanno fatto pensare che questi segmenti dovessero essere relativamente scarsi nei nostri preparati e fittamente stipati di molecole attive di RNA polimerasi. È parso anche evidente che le due grosse molecole di RNA che si trovano nei ribosomi dei procarioti vengono sintetizzate separatamente piuttosto che formarsi per frammentazione da un unico precursore, come succede negli eucarioti. I geni per questi due RNA sono stati considerati contigui e pertanto avrebbero dovuto apparire come due unità matrici di fibrille adiacenti. In base ai pesi molecolari dei due RNA dei ribosomi batterici, l'unità matrice dell'uno avrebbe dovuto essere lunga il doppio di quella dell'altro e la lunghezza totale delle due unità avrebbe dovuto avvicinarsi a 1,7 μ.

Nei nostri campioni abbiamo identificato segmenti del cromosoma di *E. coli* che corrispondono abbastanza bene a queste descrizioni. Ai segmenti sono attaccate da 60 a 70 fibrille, disposte in due unità matrici consecutive. Come nel caso delle cellule di anfibio e delle cellule HeLa, questi segmenti sono un poco piú corti di quanto non sia stato previsto ed è presumibile che ciò sia dovuto di nuovo allo srotolamento del DNA. I geni di *E. coli* per l'RNA ribosomiale non sembrano essere vicini sul cromosoma come lo sono negli eucarioti: difatti sono separati da segmenti cromosomici che mostrano l'attività poliribosomiale dei geni per la sintesi proteica. Non abbiamo ancora determinato la lunghezza minima di questi segmenti intercalari.

Il fatto che tecniche di isolamento e di preparazione piuttosto semplici ci abbiano permesso di visualizzare questi particolari strutturali dell'attività genetica in numerosi tipi di cellule ci fa sperare che la microscopia elettronica diventerà in futuro un mezzo sempre piú importante per lo studio della citogenetica. Il miglioramento delle tecniche potrà rendere possibile l'osservazione diretta dei geni attivi praticamente in qualunque tipo di cellula e, alla fine, permetterà di identificare quei geni con i loro prodotti e di sapere come funzionano in momenti diversi del ciclo vitale della cellula, come possono essere modificati variando le condizioni ambientali e come si correlano con le specifiche attività della cellula.

Le origini dei concetti di numero

Esperimenti condotti con bambini indicano che essi acquistano dapprima la nozione di numero sotto forma di sequenze ordinate e solo più tardi sotto forma di quantità. È forse l'avvio a nuovi metodi d'insegnamento

di Charles J. Brainerd

In che modo apprendiamo i concetti di numero? Questo problema è stato un argomento di discussione tra filosofi e matematici almeno dal tempo di Pitagora, nel VI secolo a.C. Oggi esso continua a essere importante, non soltanto per gli studiosi, ma anche per la società nel suo complesso; lo dimostra il trionfo della « nuova matematica » nelle scuole elementari nell'ultimo decennio. Non è però meno vero che gli studi generali sull'origine dei concetti di numero nei bambini piccoli sono relativamente rari. Recentemente ho preso parte a studi compiuti all'Università di Alberta che hanno portato all'acquisizione di nuove informazioni sull'argomento. Prima però di procedere alla descrizione di tali studi sarà utile passare in rassegna le teorie principali sull'origine dei concetti numerici.

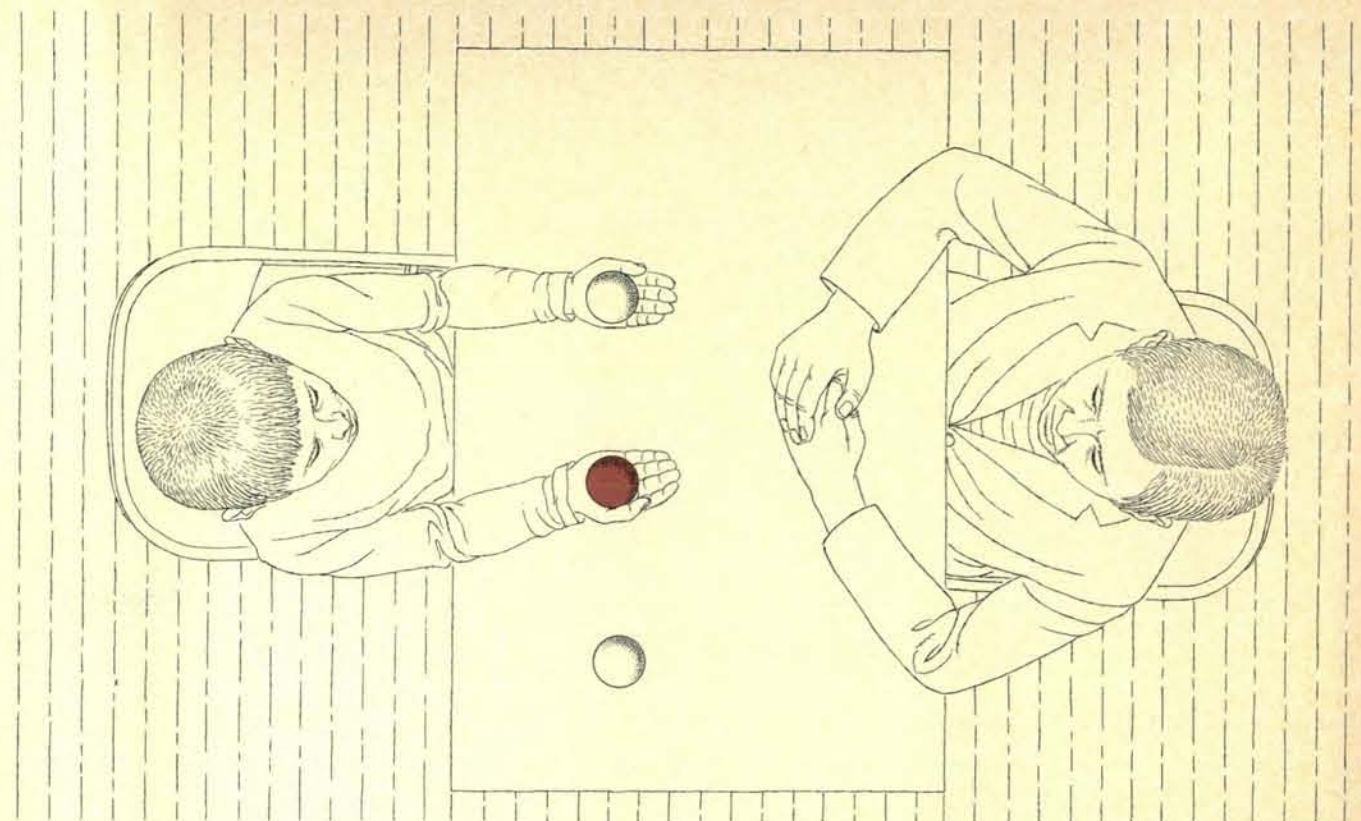
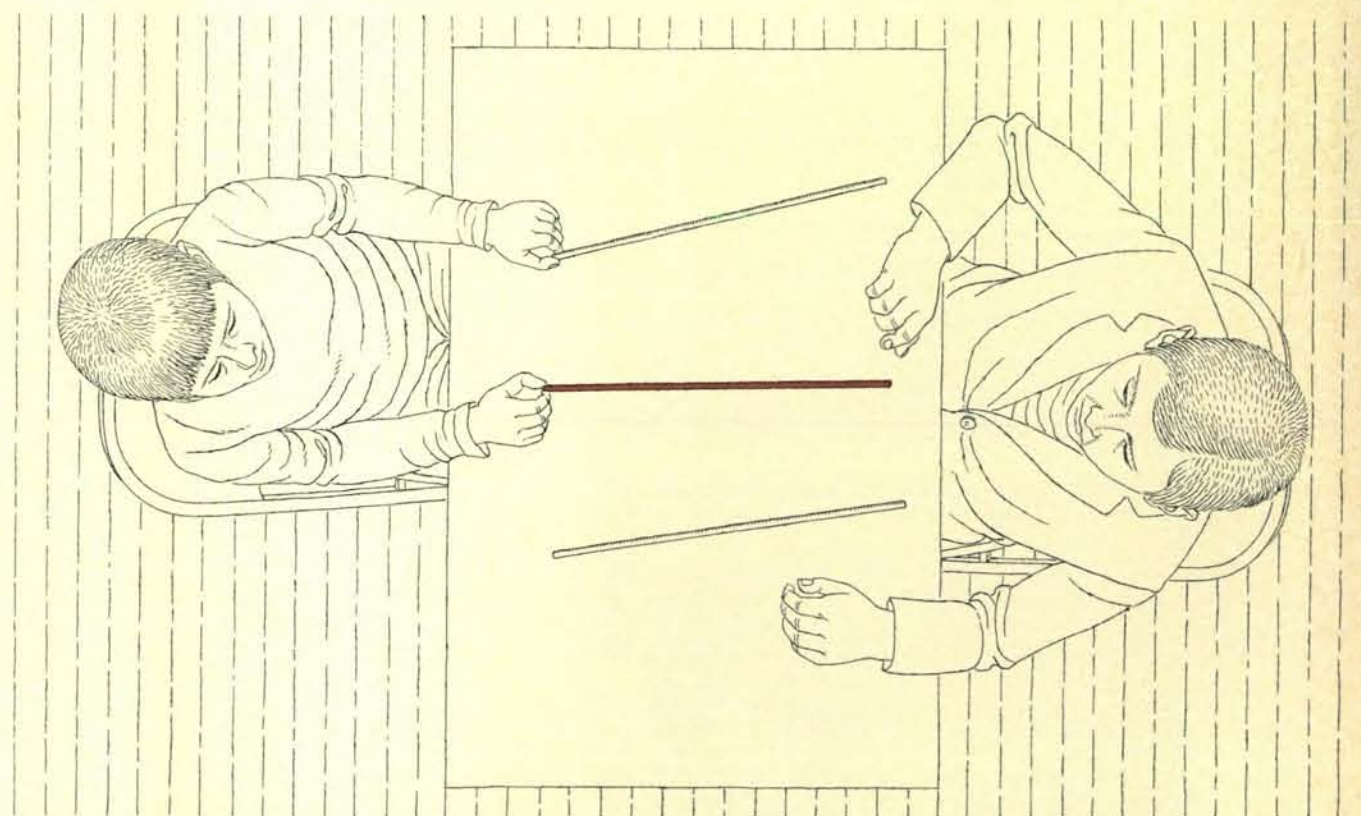
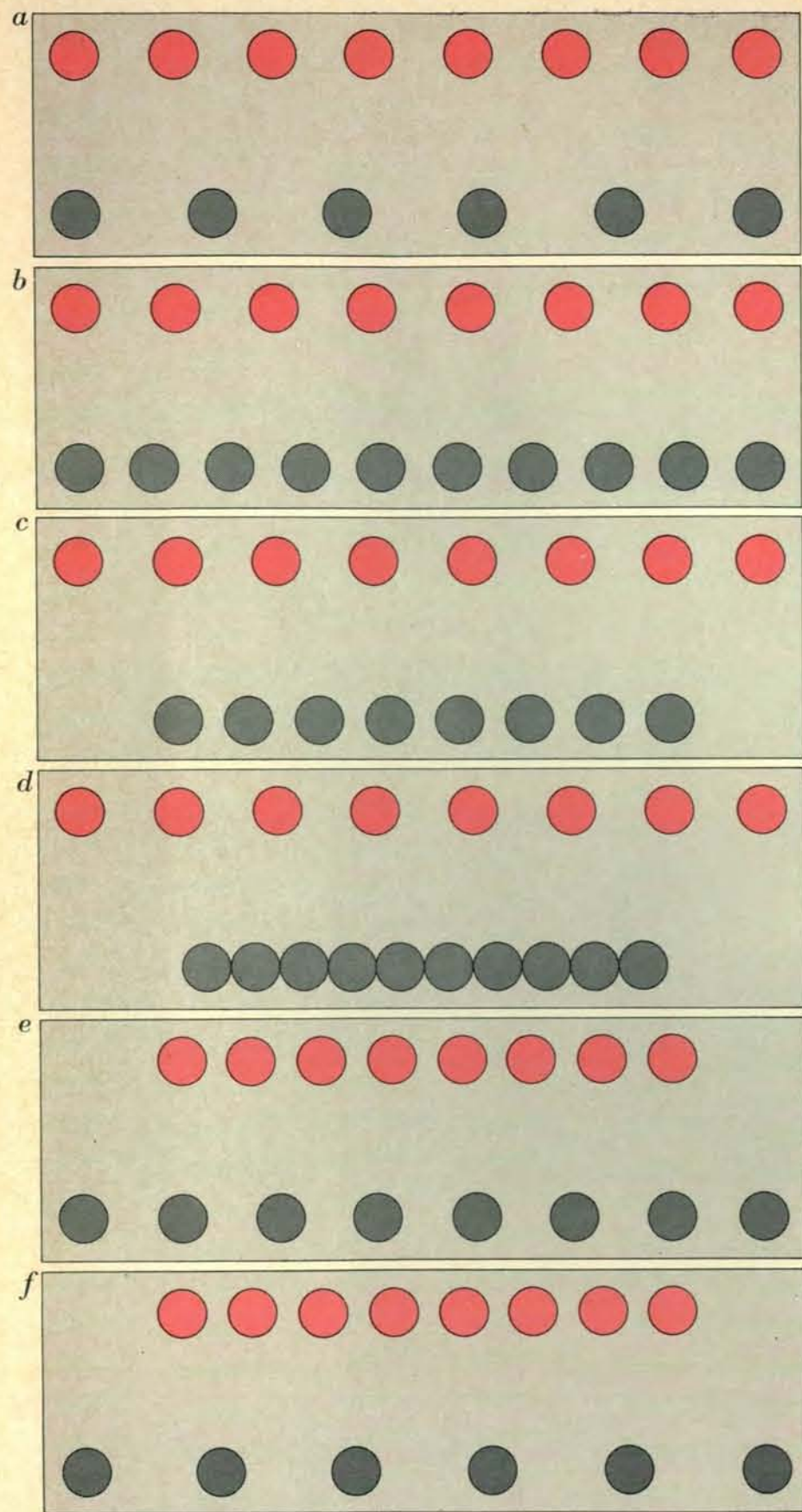
Pitagora riteneva che i cosiddetti numeri interi positivi, o numerali (1, 2, 3 e cosí via), fossero entità di origine divina e che costituissero il fondamento ultimo sia della matematica sia dell'universo. La scoperta, a opera dei pitagorici, di quantità « incommensurabili », come il rapporto tra il diametro di un cerchio e la sua circonferenza, dissiparono definitivamente la convinzione che l'universo fosse costruito sulla base dei numeri naturali. La nozione che i numeri naturali costituissero però almeno la fondazione della matematica persisté come un articolo di fede tra i matematici fino al XIX secolo inoltrato.

A quell'epoca l'atteggiamento nei confronti dei numeri naturali aveva cominciato a mutare. La centralità dei numeri naturali non fu piú considerata un fatto accettato ma una congettura che doveva essere dimostrata in modo rigoroso. Le dimostrazioni assumono normalmente la forma di una derivazione passo passo di sistemi di numeri ben noti, come i numeri razionali, i numeri reali e i numeri complessi, dai numeri naturali. Due esempi ci sono forniti dai tentativi di Karl Weierstrass e di Richard Dedekind di « aritmetizzare » l'analisi matematica. Entrambi gli studiosi derivarono i numeri reali – l'insieme di tutti i numeri razionali e irrazionali che vengono impiegati nella matematica classica – dai numeri naturali. Un terzo esempio è costituito dalla proposta di Leopold Kronecker di fondare l'intera matematica sui numeri naturali.

Altri matematici, particolarmente quelli che avevano familiarità con i contemporanei progressi che si venivano facendo nella logica simbolica, avanzarono il suggerimento che, lungi dall'essere d'origine divina, i numeri naturali fossero costruzioni della mente umana. Coloro che piú contribuirono a diffondere questo suggerimento furono Gottlob Frege, Giuseppe Peano e Bertrand Russell. Si sentiva distintamente l'esigenza di una teoria che spiegasse l'origine dei numeri naturali da qualche nozione o da un qualche complesso di nozioni piú elementare, ma in che modo una tale teoria doveva essere costruita?

Dapprima Frege, poi Peano e infine Russell si volsero alla logica simbolica come a una fonte potenziale delle nozioni fondamentali necessarie per sviluppare una teoria dei numeri naturali. Frege fu il primo dei tre a pubblicare una teoria specifica. In *Die Grundlagen der Arithmetik*, del 1884, suggerí che i numeri naturali potevano essere ridotti alla nozione di « classe » e all'opera-

La concezione cardinale determina il numero di oggetti contenuti in un insieme non attraverso un conteggio ma mediante un confronto. Gli oggetti corrispondono in numero alle cinque dita di una mano; il numero cardinale dell'insieme è 5.

La concezione ordinale determina il numero di oggetti contenuti in un insieme contando « uno, due, tre » e cosí via e non con un confronto. In questo insieme, come in tutti quelli finiti, l'ultimo numero ordinale, 5, è il numero cardinale dell'insieme.

La concezione ordinale nella forma di « piú pesante di / piú leggero di » fu verificata chiedendo ai soggetti di soppesare palle di argilla che avevano le stesse dimensioni ma peso diverso. In alcuni test la palla piú pesante era all'estremità destra della fila, in altri l'ordine fu rovesciato. In tutti i test il confronto fu fatto con una terza palla (*al centro*), di peso intermedio.

Nel secondo test sulla concezione ordinale, implicante la relazione « piú lungo di / piú corto di », ai soggetti furono presentati tre bastoncini di legno di diversa lunghezza; il piú lungo superava quello intermedio di mezzo centimetro e il piú corto ne era superato di altrettanto; il bastoncino di lunghezza intermedia veniva collocato sempre in posizione centrale fra gli altri due. Come nel test relativo al peso, l'ordine di presentazione era casuale. I soggetti confrontavano il bastoncino lungo e quello corto col bastoncino intermedio (*in colore*) ma non misero mai direttamente a confronto il bastoncino lungo con quello corto.

La concezione cardinale fu verificata chiedendo ai soggetti di indicare senza contarli, quali dei sei insiemi di pallini (*a-f*) fossero in corrispondenza biunivoca e quali in una corrispondenza non biunivoca. La fila superiore e quella inferiore di pallini in ogni insieme avevano diverso colore. Se entrambe le file contenevano lo stesso numero di pallini, questi venivano spaziati in modo che le file non avessero uguale lunghezza. Quando i pallini erano in numero diverso venivano spaziati in modo tale che le file avessero ugual lunghezza o che fosse piú lunga la fila contenente meno pallini.

zione di « corrispondenza » in virtú della quale le classi vengono quantificate. Secondo Frege, ogni numero naturale *n* dev'essere considerato una « classe sovraordinata » ciascuno dei cui membri, le « classi subordinate », contiene *n* elementi. Date due classi subordinate *A* e *B*, esse vengono dette membri della stessa classe sovraordinata, ossia esempi dello stesso numero, se e solo se una corrispondenza biunivoca può essere istituita tra i loro rispettivi elementi. Se invece la corrispondenza non è biunivoca, si dice che *A* e *B* sono esempi di numeri diversi.

Sostanzialmente la teoria di Frege afferma che la serie dei numeri naturali presenta un problema generale di quantificazione ma che esso può essere ristretto alla nozione piú limitata della concezione cardinale o delle classi quantificanti. L'esempio piú comune della concezione cardinale ci è dato dal confronto di oggetti.

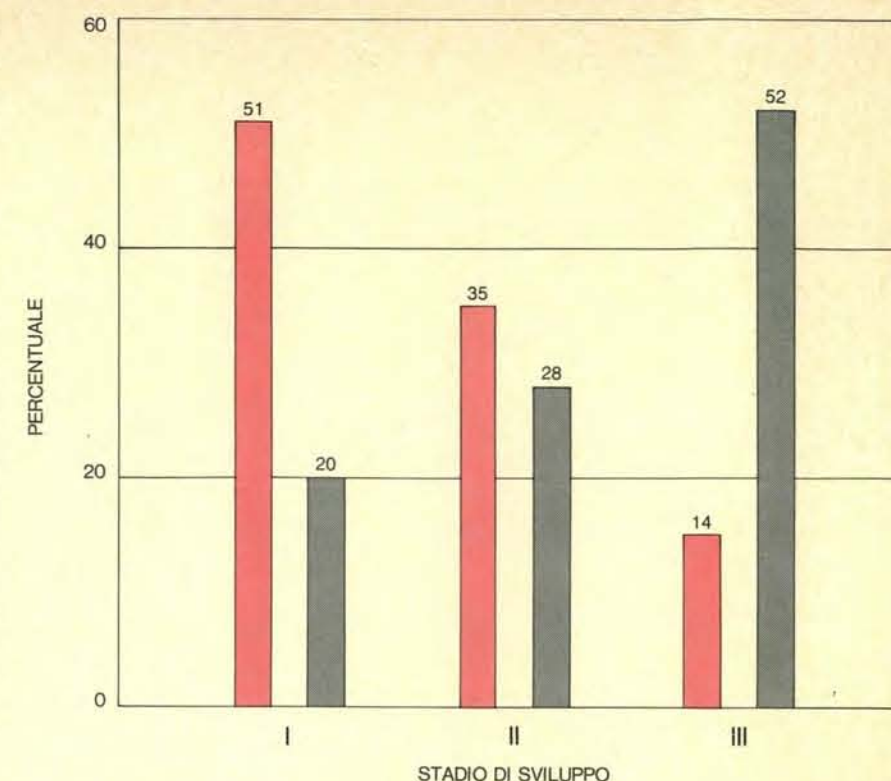
La teoria dei numeri cardinali di Frege rimase pressoché sconosciuta finché Russell non la riscoprí nel 1901. Russell pubblicò successivamente la teoria, riconoscendo pienamente la priorità di Frege, nei *Principles of Mathematics* (1903), nei *Principia mathematica* (1910-1913), scritti insieme ad Alfred North Whitehead, e nell'*Introduction to Mathematical Philosophy* (1919).

Fra l'epoca in cui Frege pubblicò per la prima volta la teoria cardinale e quella in cui essa fu riscoperta da Russell, Peano sviluppò una seconda teoria sui numeri naturali. Questa teoria apparve per la prima volta nella sua opera *Formulaire de mathématique* (1894) nella forma di cinque assiomi che ripeterò qui leggermente modificati. I. L'unità è un numero naturale. II. Il successivo di un numero naturale è un numero naturale. III. Due numeri naturali diversi non hanno lo stesso successivo. IV. Il numero naturale 1 non è il successivo di alcun numero naturale. V. Se una serie contiene sia il numero 1 sia il successivo di ogni numero naturale, allora la serie contiene tutti i numeri naturali.

In sostanza la teoria di Peano colloca i numeri naturali in una relazione ordinale o, nel linguaggio della logica simbolica, in una « relazione transitiva asimmetrica ». Se siamo disposti ad accordarci sul fatto che la relazione *R* che si ottiene tra ogni coppia non identica di numeri naturali sia una relazione ordinale, allora la serie completa dei numeri naturali potrà essere costruita un passo dopo l'altro con l'aiuto della legge dell'induzione matematica. Come la teoria cardinale di Frege, la teoria di Peano asserisce che le serie di numeri naturali presentano un problema generale di quantificazione. A differenza della teoria di Frege, la teoria ordinale di Peano riduce però il problema generale alla nozione piú ristretta di relazioni quantificanti transitive asimmetriche, ossia alla concezione ordinale. L'esempio piú comune è il conteggio di oggetti.

Quale delle due teorie, quella cardinale o quella ordinale, sia matematicamente preferibile è un problema che non ha mai trovato una risposta soddisfacente per tutti. Obiezioni ragionevoli possono essere sollevate nei confronti di entrambe. La teoria cardinale, per esempio è soggetta al famoso paradosso, scoperto nel 1901 da Russell, relativo alla classe composta da tutte quelle classi che non contengono se stesse come elemento. In breve, se *Y* è la classe composta da tutte quelle classi che non contengono se stesse come elemento, allora si può dimostrare che *Y* contiene e non contiene se stessa come elemento.

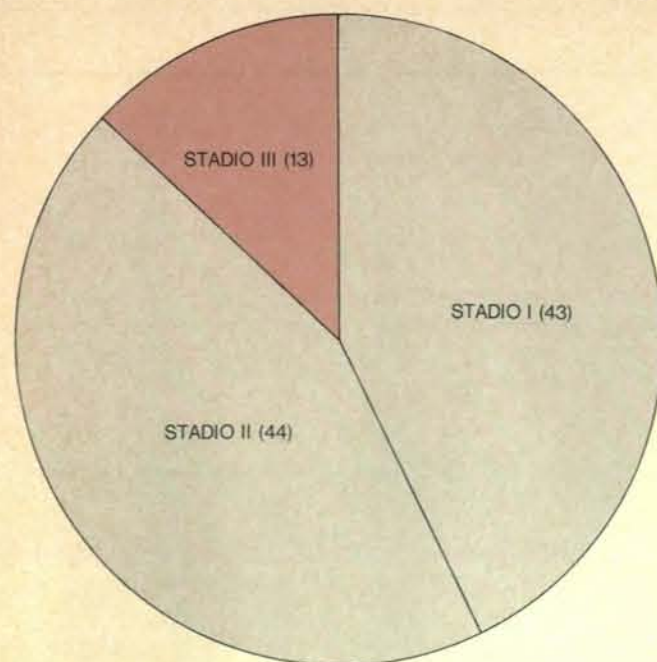
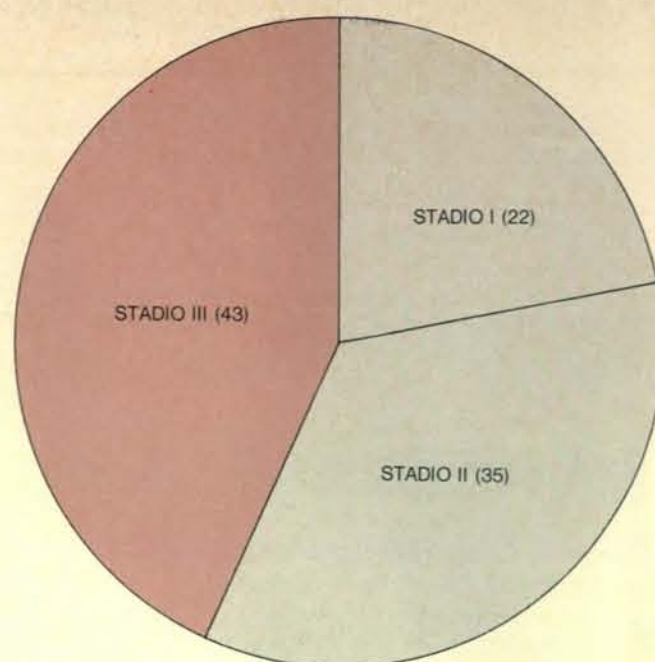
Quanto alla teoria ordinale, come sottolineò Russell, se i cinque assiomi di Peano sono chiaramente soddisfatti della serie dei numeri naturali, essi sono soddisfatti anche da altri sistemi di numeri. Per esempio, saranno soddisfatti dalle frazioni razionali (1, 1/2, 1/3, 1/4 e cosí via), come pure da ogni serie di entità matematiche o empiriche che abbia un inizio, che non presenti ripetizioni, che non abbia una fine e che sia tale che ogni entità possa essere raggiunta mediante un numero finito di passaggi. In breve, il dominio d'applicazione della teoria ordinale è molto piú ampio della serie dei numeri naturali.

Non esistendo una base matematica universalmente accettata per scegliere fra la teoria cardinale e quella ordinale, la scelta diventa una questione soggettiva. Tipicamente, essa viene a essere determinata dal grado di simpatia che si prova per l'una o l'altra di tre scuole moderne di pensiero matematico: il logicismo, il formalismo e l'intuizionismo. Coloro che inclinano verso il logicismo dimostrano una preferenza per la teoria cardinale, scelta abbastanza naturale se si considera che i padri della teoria, Frege e Russell, sono i principali fondatori del logicismo. I simpatizzanti per il formalismo propendono verso la teoria ordinale; questa preferenza è spiegata probabilmente dal fatto che gli assiomi di Peano sembrano privare il concetto di numero di un « significato » intrinseco. Quanto agli intuizionisti, essi sono tornati in effetti alla posizione pitagorica secondo cui

La comprensione dei concetti, qual era rivelata dai risultati dei test, consentí di collocare ciascuno dei 180 soggetti in uno di tre distinti stadi di sviluppo. I bambini appartenenti al I stadio erano totalmente incapaci di afferrare il principio cardinale (*in colore*) o quello ordinale (*in grigio*). Quelli appartenenti al II stadio erano in grado di stabilire una corrispondenza non biunivoca (*in colore*) e una successione da sinistra a destra (*in grigio*) ma non riuscivano ad afferrare la corrispondenza biunivoca e relazioni da destra a sinistra. I bambini appartenenti al III stadio compresero pienamente il concetto cardinale (*in colore*) e quello ordinale (*in grigio*). Come indica la percentuale di bambini appartenenti ai vari stadi, la concezione ordinale del numero appare nel pensiero dei bambini molto prima della concezione cardinale del numero.

i numeri naturali devono essere accettati senza ulteriore analisi come il fondamento della matematica. Essi negano che i numeri naturali siano un'invenzione di menti matematiche e suggeriscono una tesi « psicologica »: la serie dei numeri naturali è un'intuizione innata, presente alla nascita in tutti i membri normali della specie umana.

Gli studiosi non matematici tendono a considerare con profonda indifferenza le torture che i matematici soffrono affrontando problemi fondamentali come la natura dei numeri. Essi hanno imparato da secoli di esperienze che il fatto che le fondazioni di un sistema o di un concetto matematico non siano sicure non deve dissuadere dall'usare il sistema nel proprio lavoro. Al contrario, le nozioni matematiche le cui fondazioni sono state oggetto di discussioni continue hanno dato spesso i risultati migliori: la nozione di infinitesimo è forse l'esempio piú noto.

Un infinitesimo può essere definito una quantità che, pur essendo infinitamente piccola, è nondimeno maggiore di zero. Il calcolo infinitesimale, quale fu formulato originariamente da Newton e da Leibniz, implicava l'uso di infinitesimi; Newton li chiamò flussioni. Punti di vista violentemente contrastanti contrassegnarono le successive discussioni del problema se un'entità curiosa come un infinitesimo possa essere considerata un concetto matematico legittimo. Nella seconda metà del XIX secolo la discussione parve finalmente sopita dal successo conseguito da Weierstrass nel riformulare il « calcolo » senza gli infinitesimi. Nel passato decennio, però, l'opera di Abraham Robinson sull'analisi non-standard ha resuscitato gli infinitesimi (si veda l'articolo *L'analisi non-standard* di Martin Davis e Reuben Hersh, in « Le Scienze », n. 49, settembre 1972). Per tutto quell'intervallo di tempo il calcolo infinitesimale è stato nondimeno regolarmente applicato nelle scienze fisiche e biologiche senza eccessive preoccupazioni per le sue oscure basi matematiche e senza che ciò comportasse danno per le discipline che se ne sono servite.

Ma il fatto che né le scienze né la società abbiano sofferto danni dall'in-

Un netto contrasto tra la comprensione elementare della nozione cardinale e di quella ordinale di numero risulta evidente da questi due grafici. Su 93 bambini che afferrarono pienamente la concezione ordinale del numero (*a sinistra*), soltanto il 13 per cento (*in colore*) possedeva anche una comprensione piena della concezione cardinale. La concezione cardinale non fu compresa affatto da 92 bambini (*a destra*); il 43 per cento di essi (*in colore*) aveva però una buona comprensione del concetto ordinale.

differenza (o piú precisamente dall'ignoranza) degli studiosi nei confronti della discussione sugli infinitesimi, non dovrebbe indurci a supporre che la discussione sull'origine del concetto di numero possa essere ignorata con uguale sicurezza di non doverne subire grave danno. A differenza dell'infinitesimo, il numero non è una proprietà esclusiva del matematico, né è in gran parte una sua proprietà personale. Il numero è stato un concetto di importanza sociale fin dagli inizi della storia scritta. L'importanza per la società è cresciuta enormemente col sorgere della civiltà industriale.

Nella maggior parte delle nazioni occidentali oggi i bambini vengono posti di fronte al concetto di numero fin da quando imparano a parlare. Piú o meno casuale in principio, questa stimolazione diventa piú intensa e al tempo stesso piú sistematica con l'inizio dell'istruzione formale. Nei primi anni della scuola elementare il 50 per cento circa del programma è costituito normalmente dall'insegnamento dei numeri naturali e dei metodi per operare con essi. Il bambino impara a sommare, sottrarre, moltiplicare e dividere i numeri naturali e infine a derivare da essi altri sistemi di numeri. Noi ci aspettiamo che i nostri figli posseggano, una volta arrivati al periodo della pubertà, una competenza numerica molto superiore a quella di un greco o di un romano adulto istruito di duemila anni fa. Siamo addirittura arrivati al punto di sviluppare una serie di etichette che implicano una turpitudine mentale da parte di quei bambini, per altro normali, che non riescono a raggiungere quei livelli di capacità matematica che consideriamo desiderabili; parliamo allora di « incapacità di apprendimento » e di « insufficienza ».

Considerando l'accento che la nostra società pone sull'abilità numerica, ci si potrebbe attendere che sin da molto tempo fa siano state raccolte estese informazioni su come il concetto di numero si chiarisca naturalmente nel corso dello sviluppo mentale di un bambino. La verità è che possediamo pochissimi fatti certi sull'emergere di nozioni numeriche nel pensiero del bambino. È proprio la scarsità di tali dati a conferire rilevanza sociale alla discussione matematica sulle origini del concetto di numero che abbiamo appena delineato.

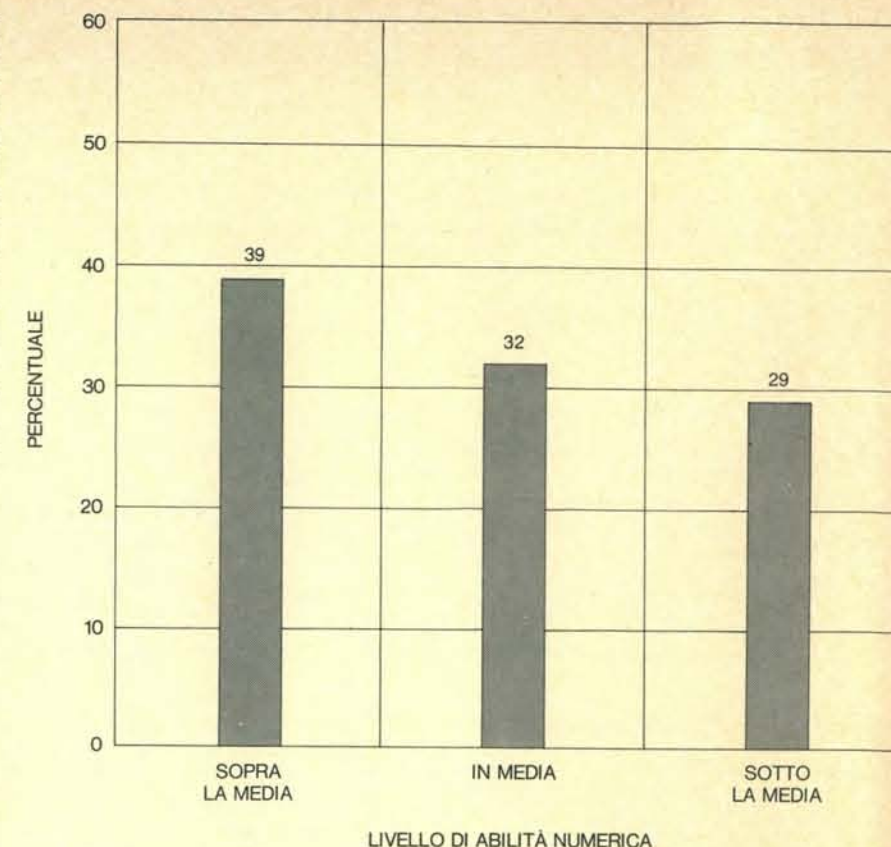
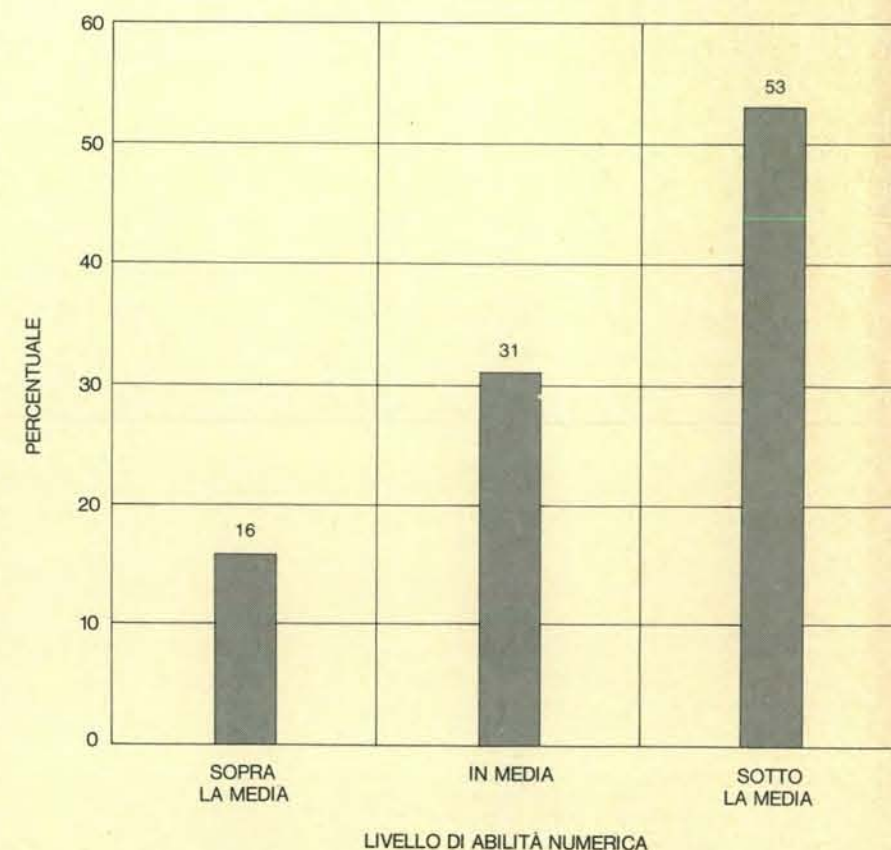
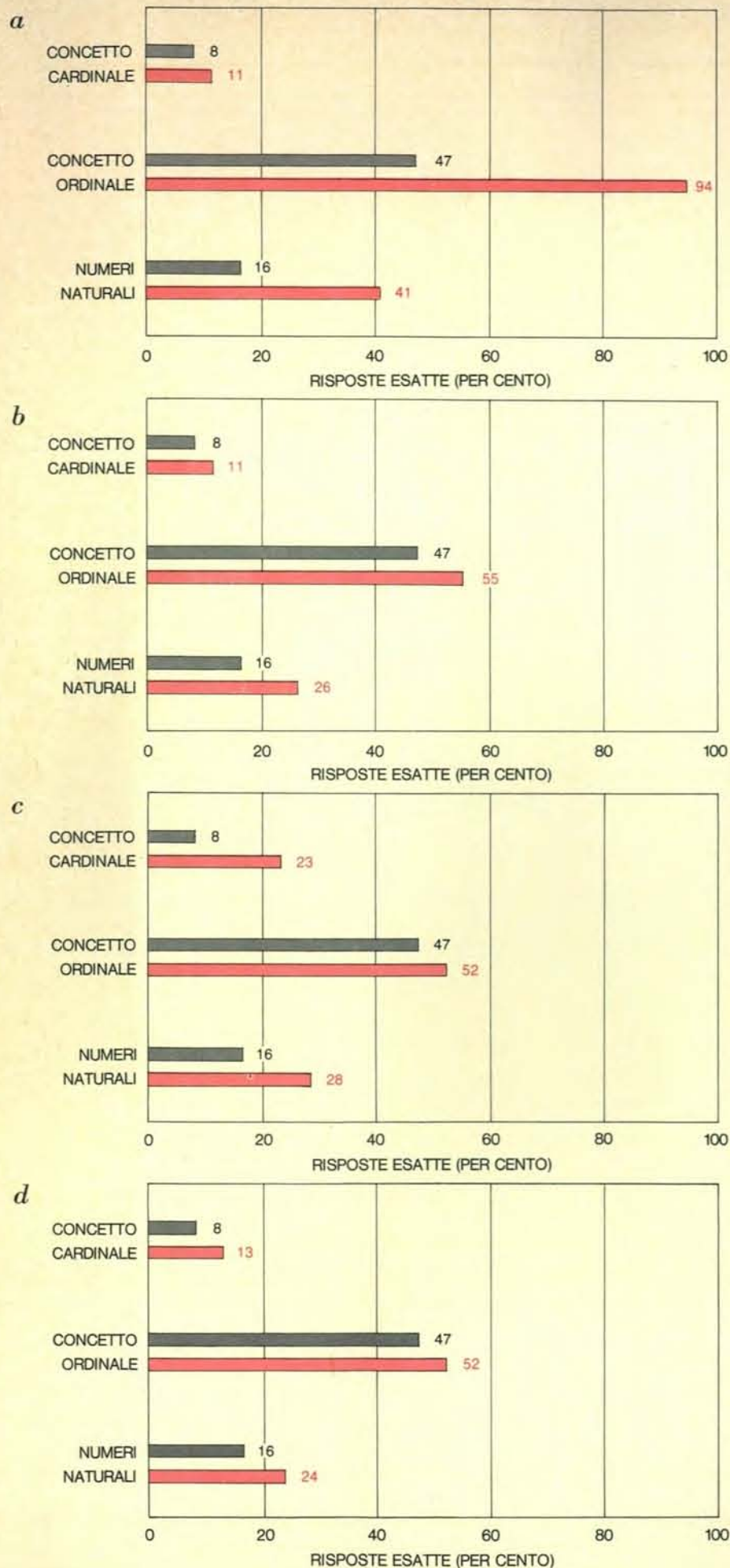
Quando il matematico dice che un concetto *Y* (per esempio un numero naturale) « si fonda » o « si basa » su un altro concetto *X* (per esempio la concezione cardinale o quella ordinale), egli intende che in senso logico *X* sia piú fondamentale di *Y*. Quando lo psicologo dello sviluppo fa la stessa asserzione, intende dire che *X* svolge una parte necessaria nell'emergere di *Y* nel pensiero del bambino che sta crescendo. Dal punto di vista dello psicologo, tale asserzione conduce a due predizioni principali, che possono essere verificate sperimentalmente. La prima predizione è che *X* apparirà invariabilmente prima di *Y* nel corso dello sviluppo mentale del bambino. La seconda è che progressi in *Y* saranno una conseguenza di progressi in *X*.

Strategie di ricerca diverse sono usate per esaminare queste due predizioni. La predizione « *X* prima di *Y* » richiede studi normativi, ovvero sui risultati medi. Lo studio normativo consta di quattro fasi. In primo luogo vengono progettati reattivi mentali o test che misurano la presenza dei concetti in questione nel pensiero del bambino. In secondo luogo, i bambini che devono essere sottoposti al test vengono tratti dalle classi di età in cui i concetti in questione vengono normalmente acquisiti. Il terzo stadio consiste nel verificare la presenza o l'assenza dei concetti. Il quarto è il confronto tra i risultati dei test al fine di determinare la presenza o l'assenza di correlazioni interconcettuali del tipo predetto « *X* prima di *Y* ».

La predizione « meglio *X* significa meglio *Y* » è verificata dal cosiddetto esperimento di transfer, un metodo leggermente piú complicato del metodo normativo. Esso comprende sette fasi; le prime tre sono uguali alle prime tre fasi dello studio normativo. Considerate globalmente, esse costituiscono la fase preliminare dell'esperimento di transfer. Le informazioni raccolte in questa fase vengono poi usate per suddividere i soggetti in gruppi sperimentali e in gruppi di controllo. Nella quinta fase i gruppi sperimentali ricevono un addestramento inteso a migliorare le loro prestazioni nell'uno o nell'altro dei concetti in questione. Dopo di ciò i test applicati nella fase preliminare vengono assegnati ai gruppi sperimentali e ai gruppi di controllo. La settima e ultima fase consiste nel confronto dei risultati dei test al fine di determinare il grado di progresso legato all'addestramento e di stabilire se l'addestramento a un concetto specifico tenda o no a « trasferire » e migliorare le prestazioni rispetto a un altro concetto per il quale non era stato somministrato alcun addestramento.

Sia gli esperimenti normativi sia quelli di transfer appaiono semplici, ma una trappola seria attende gli investigatori incauti. Nel costruire i test destinati a consentire di stimare la comprensione dei vari concetti da parte del soggetto bisogna guardarsi da metodi di valutazione definiti in modo inesatto che comportino solo una somiglianza vaga o intuitiva nei confronti dei concetti che devono essere determinati quantitativamente. Nel caso in esame, per esempio, era essenziale che i test concernenti i concetti cardinale e ordinale fossero dedotti esplicitamente dalla concezione matematica di queste nozioni. Se non si sarà seguito questo procedimento, i risultati degli esperimenti potranno avere solo un valore di suggerimento, ma non un valore conclusivo. Anche se la trappola può sembrare manifesta, non fu evitata nella maggior parte delle investigazioni compiute in passato sullo sviluppo dei concetti di numero nei bambini, non esclusi i classici studi compiuti da Jean Piaget attorno al 1940.

All'Università di Alberta intraprendemmo recentemente due studi su vasta scala sullo sviluppo di tre elementi aritmetici nel pensiero di bambini piccoli: il concetto cardinale (il « confronto di insiemi » di Frege e Russell), il concetto originale (la quantificazione sequenziale di Peano) e la capacità dei bambini di operare con numeri naturali. Gli studi si proponevano di rispondere a tre quesiti legati fra loro del tipo « *X* prima di *Y* »: Quale dei due sorge prima, il concetto cardinale o quello ordinale? Che cosa si presenta prima, il concetto ordinale o la conoscenza dei numeri naturali? E infine viene prima il concetto cardinale o la conoscenza dei numeri naturali?

Ci occupammo dapprima esclusivamente del problema della priorità tra il concetto cardinale e quello ordinale. I soggetti esaminati furono 180 bambi-

Test condotti su altri 180 bambini delle elementari dimostrarono che esiste un'associazione tra la comprensione del concetto ordinale e la capacità di operare con i numeri. Su 119 bambini appartenenti al III stadio nella comprensione del concetto ordinale, tre quarti circa si dimostrarono al di sopra della media o in media nell'abilità numerica.

Test condotti sugli stessi bambini dimostrarono che non esiste un'associazione tra la comprensione del concetto cardinale e la capacità di operare con i numeri. Dei 95 bambini appartenenti al I stadio nella comprensione del concetto cardinale di numero, quasi la metà dimostrarono un'abilità numerica superiore alla media o pari alla media.

ni canadesi di età compresa tra i cinque e i sette anni, l'intervallo di età in cui il bambino medio acquista per la prima volta i concetti cardinale e ordinale e in cui sviluppa una qualche abilità nel servirsi dei numeri naturali. I test relativi al principio ordinale servivano a valutare la capacità del bambino di quantificare due familiari relazioni transitive asimmetriche: « piú pesante di/piú leggero di » e « piú lungo di/piú corto di ».

Si cominciava col presentare a ogni bambino tre palle d'argilla disposte in fila su un tavolo. Le tre palle avevano le stesse dimensioni; le due palle all'estremità della fila avevano lo stesso colore (*si veda la figura a pagina 85 in alto*). Pur avendo le stesse dimensioni, le palle avevano peso diverso. In alcuni test la palla piú pesante veniva collocata all'estremità sinistra della fila e quella piú leggera all'estremità destra; in altri test l'ordine veniva rovesciato. Al bambino si chiedeva di confrontare il peso della palla di mezzo e di quella di sinistra, e poi il peso della palla di mezzo e di quella di destra. Una volta eseguite queste operazioni, gli si chiedeva quale relazione quantitativa intercorresse tra le palle collocate ai due estremi della serie. Al bambino non era stata data alcuna informazione su questo tipo di relazione, cosicché per rispondere correttamente alla domanda, doveva avere una comprensione intuitiva del principio ordinale.

Esperimenti di transfer (*a sinistra*) verificarono se un addestramento all'uso del principio cardinale e di quello ordinale avrebbero influito sulla capacità dei bambini di operare con numeri naturali. Duecentoquaranta bambini prescelti per questi test furono divisi in quattro gruppi di 60, curando che la loro composizione fosse omogenea dal punto di vista della comprensione del concetto cardinale, di quello ordinale e dell'abilità numerica (*linee grige*). I gruppi furono poi accoppiati due a due; i bambini appartenenti al gruppo *a* e quelli del gruppo *c* furono particolarmente addestrati, rispettivamente, nella concezione ordinale e in quella cardinale, mentre i bambini dei gruppi *b* e *d*, considerati gruppi di controllo, furono semplicemente sottoposti a test di routine. Dopo otto sedute di addestramento o di semplici test, fu assegnato a tutt'e quattro i gruppi un test finale (*linee in colore*). L'addestramento nella concezione ordinale raddoppiò quasi i risultati dei bambini del gruppo *a*; un effetto di transfer è evidente anche nei miglioramenti realizzati dai bambini del gruppo *a* nell'abilità numerica. L'addestramento nella concezione cardinale triplicò quasi i risultati dei bambini del gruppo *c*, i quali ebbero però risultati modesti nell'abilità numerica, sulla quale non ebbe alcun effetto significativo l'addestramento nella concezione cardinale.

In un secondo test venivano presentati al bambino tre bastoncini disposti in una serie da sinistra a destra. I tre bastoncini sembravano della stessa lunghezza, ma in realtà quelli disposti agli estremi della serie, i quali erano dipinti nello stesso colore, erano rispettivamente piú lungo e piú corto di mezzo centimetro rispetto al bastoncino centrale. Come nel caso delle palle d'argilla, il bastoncino piú corto fu collocato talvolta a sinistra e talvolta a destra (*si veda la figura a pagina 85 in basso*). Al bambino veniva chiesto di confrontare il bastoncino di mezzo con quello di sinistra e poi il bastoncino di mezzo con quello di destra; quindi gli veniva chiesto quale relazione quantitativa esistesse tra i bastoncini che si trovavano ai due estremi della serie. Anche qui, per rispondere in modo corretto alla domanda, era necessaria una nozione del principio ordinale.

I due test sul principio ordinale furono assegnati ai 180 bambini. Fu subito chiaro che il gruppo comprendeva bambini a tre distinti stadi di sviluppo. I bambini appartenenti a quello che chiamerò stadio I erano totalmente incapaci di stabilire una successione. Quelli appartenenti al II erano capaci di un ordinamento spaziale ma non di un vero ordinamento quantitativo. Intendo dire che essi rispondevano con precisione quando l'ordine dal piú leggero al piú pesante era da sinistra a destra ma non quando l'ordine veniva rovesciato. I bambini appartenenti al III stadio avevano una chiara nozione del concetto ordinale e rispondevano in modo esatto quale che fosse l'ordine di presentazione degli oggetti.

Poiché il principio cardinale consiste nella quantificazione di classi sulla base della corrispondenza tra le rispettive classi, il test relativo si proponeva di valutare la capacità del bambino di quantificare coppie di classi che presentassero una corrispondenza biunivoca o non biunivoca. Il test era composto da sei problemi (*si veda la figura a pagina 86*). In ciascun problema venivano presentate al bambino due file parallele di pallini; i pallini della fila superiore erano rossi, quelli della fila inferiore blu. Il numero minimo di pallini presenti in ogni fila era sei e il numero massimo dieci. Quando una fila aveva lo stesso numero di pallini dell'altra (corrispondenza biunivoca), i pallini venivano spaziati in modo tale che le due file avessero nondimeno una lunghezza disuguale. Analogamente, quando una fila aveva piú o meno pallini dell'altra (corrispondenza non biunivoca), i pallini erano spaziati in modo tale che le due file presentassero la stessa lunghezza o che la fila con meno pallini fosse la piú lunga. Questo accorgimento era destinato a impedire che il bambino formasse il suo giudizio sulla base di indizi percettivi grossolani. Per assicurare inoltre che le sei coppie fossero quantificate solo attraverso la corrispondenza, fu data al bambino l'istruzione di non contare i pallini. (Il conteggio è, ovviamente, un'operazione ordinale.)

Tutti e 180 i bambini furono sottoposti al test sul principio cardinale. Come nel caso del test sul principio ordinale, si riscontrarono tre stadi di sviluppo. I bambini al I stadio erano totalmente incapaci di afferrare il principio della corrispondenza. I bambini al II stadio si dimostrarono in grado di quantificare le corrispondenze piú di uno a uno ma non le corrispondenze uno a uno. I bambini al III stadio furono in grado di applicare il principio cardinale sulla base di entrambi i tipi di corrispondenza.

La nostra scoperta piú importante scaturí dal confronto tra i risultati dei due test. Già la differenza nella distribuzione dei bambini in ciascuno dei tre stadi di sviluppo aveva suggerito che il principio ordinale comparisse nel pensiero dei bambini molto tempo prima del principio cardinale. Per esempio, nel test sul concetto ordinale, 93 bambini su 180 risultarono appartenere al III stadio e 24 al I, mentre nel test sul concetto cardinale solo 15 bambini risultarono appartenere al III stadio e ben 92 al I (*si veda la figura a pagina 87*). Quando si confrontarono i risultati ottenuti da uno stesso bambino nei due test, la posizione del principio ordinale « *X* prima di *Y* » divenne ancor piú evidente. Solo 12 dei 93 bambini che risultarono appartenere al III stadio nel test relativo al principio ordinale operarono al livello del III stadio anche nel test sul principio cardinale. Dei 92 bambini che furono assegnati al I stadio nel test cardinale, 40 salirono al livello del III stadio nel test relativo al principio ordinale e altri 32 a quello del II stadio (*si veda la figura a pagina 88*). Tutto sommato non si potrebbe chiedere una risposta piú chiara al quesito se il primo concetto ad apparire nel pensiero del bambino sia quello ordinale o quello cardinale.

Per il nostro test normativo scegliemmo altri 180 bambini, che dividemmo in gruppi uguali di bambini che andavano all'asilo (da cinque a sei anni di età) e di scolari di prima elementare (da sei a sette anni d'età). Ai bambini furono assegnati i test sul principio ordinale e su quello cardinale descritti sopra e inoltre un test diviso in due parti sull'abilità nell'uso dei numeri naturali. La prima parte di quest'ultimo test serviva a valutare la capacità dei bambini a sommare i primi quattro numeri naturali; esso constava dei 16 problemi $1 + 1 = ?$, $1 + 2 = ?$ fino a $4 + 4 = ?$ La seconda parte del test mirava a determinare la capacità di eseguire sottrazioni da 8 o da numeri inferiori ottenendo differenze corrispondenti a uno dei primi quattro numeri naturali (l'inverso dell'operazione precedente). Anch'essa constava di 16 problemi: $2 - 1 = ?$, $3 - 1 = ?$, fino a $8 - 4 = ?$ Ai 90 bambini di età compresa fra i sei e sette anni furono assegnate entrambe le parti del test, mentre solo la prima parte fu sottoposta ai 90 bambini di età compresa fra i cinque e i sei anni, i quali non avevano ancora imparato la sottrazione all'asilo.

Le singole domande furono presentate simultaneamente in forma verbale e scritta. Per esempio, per il primo problema fu distribuito ai bambini un foglio di carta su cui era stampato $1 + 1 = ?$ e fu loro chiesto contemporaneamente: « Quante mele sono una mela e una mela? » Se il bambino dava la risposta esatta per iscritto o a voce, o in entrambe le forme, il problema veniva considerato risolto e come tale registrato.

Dopo aver raccolto i risultati dei test, ci consultammo con ispettori, direttori e maestri di varie scuole elementari al fine di determinare quali capacità di operare con i numeri si dovessero considerare inferiori alla media, quali si dovessero considerare corrispondenti alla media e quali superiori in bambini di quelle età. Sulla base di queste consultazioni stabilimmo tre livelli di prestazione. Da 12 a 16 risposte esatte equivalevano a una prestazione superiore, da 6 a 11 al livello medio e da zero a 5 a un risultato inferiore alla media. In base ai risultati ottenuti dividemmo dunque i soggetti in tre livelli di abilità numerica.

L'analisi di questo studio rivelò innanzitutto che i risultati dei test confermavano le nostre precedenti scoperte sul principio ordinale e cardinale. Nel secondo gruppo di 180 bambini il principio ordinale emerse secondo gli stessi tre stadi in cui si era presentato nel primo gruppo e la stessa corrispondenza si trovò anche per la comparsa della concezione cardinale; analogamente, la concezione ordinale emerse molto tempo prima di quella cardinale. Il problema esaminato successivamente fu se la concezione ordinale fosse anteriore o posteriore all'abilità di eseguire operazioni con i numeri naturali. Per trovare la risposta stabilim-

mo una comparazione tra i vari stadi di sviluppo del bambino (stadio I, II o III) per quanto concerneva l'emergere della concezione ordinale del numero e i gradi di abilità con i numeri (inferiore alla media, media o superiore alla media). Dei 180 bambini 119 risultarono appartenere, quanto all'emergere del concetto ordinale di numero, al III stadio, ossia allo stadio piú alto, 39 al II stadio e 22 al I stadio. Dei 119 bambini al III stadio, 46 presentarono il livello massimo di abilità con i numeri, 38 un'abilità media e 35 rimasero al di sotto della media. Il confronto dimostrò, con nostra soddisfazione, che la concezione ordinale precede naturalmente l'abilità con i numeri.

Restava una terza questione: che cosa compare prima tra l'abilità con i numeri e la concezione cardinale? Confrontammo allora i risultati ottenuti con i test sul principio cardinale e quelli relativi al grado di abilità con i numeri. Dei 180 bambini, soltanto 33 appartenevano al III stadio, ossia al livello massimo, nella comprensione del principio cardinale, 38 appartenevano al II e 95 al I. Di questi 95, 15 rivelarono un'abilità numerica superiore alla media e 30 un'abilità media. I restanti 50 rivelarono il livello piú basso di abilità con i numeri. Il confronto ci permise di stabilire che l'abilità nell'operare con i numeri naturali precede l'emergere della nozione cardinale di numero.

Complessivamente, i nostri due studi normativi presentarono una sequenza invariante nello sviluppo del concetto di numero nel bambino. La concezione ordinale era la prima a emergere, seguita dalla capacità di operare con i numeri naturali e poi dalla concezione cardinale. Questa sequenza rende assai verosimile l'ipotesi che la capacità di operare con i numeri naturali si fondi su una previa comprensione del concetto ordinale di numero e non del concetto cardinale.

È possibile interpretare la sequenza stabilita dai nostri test normativi in altri modi. Per esempio, si potrebbe asserire che il principio ordinale, la capacità di operare con numeri naturali e il principio cardinale vengano acquisiti in modo relativamente indipendente l'uno dall'altro e che la sequenza osservata non implichi necessariamente una dipendenza cognitiva tra questi concetti. Fu questa la ragione che ci indusse a progettare il nostro terzo esperimento: l'esperimento di transfer, destinato a verificare la validità della nostra previsione « meglio X significa meglio Y ». I nostri soggetti furono 240 bambini di età compresa tra cinque e sei anni; ciascun bambino fu esaminato una volta la settimana per un periodo di 10 settimane. Durante il primo incontro furono sottoposti al bambino i tre test usati negli studi normativi precedenti. I bambini furono allora suddivisi in quattro gruppi di 60 bambini ciascuno. Si fece in modo, scegliendo accuratamente i bambini, che i gruppi avessero lo stesso livello di partenza, per esser certi che differenze destinate a emergere eventualmente piú tardi nei risultati non fossero dovute a differenze iniziali già esistenti nei gruppi.

Furono poi scelti due gruppi, uno dei quali designato quale gruppo sperimentale e l'altro come gruppo di controllo per l'addestramento nella concezione ordinale, mentre gli altri due gruppi furono designati come gruppo sperimentale e gruppo di controllo per l'addestramento nella concezione cardinale. Nelle successive otto settimane, ai due gruppi sperimentali fu assegnato un test settimanale del tipo usato negli studi normativi. L'addestramento fu rinforzato da un semplice effetto di retroazione: si ottenne questo risultato comunicando semplicemente a ogni bambino se le risposte da lui date alle domande del test erano esatte o no; in ciascuna delle otto sedute ci furono approssimativamente 15 minuti di insegnamento col sistema della retroazione che abbiamo appena spiegato. Ai gruppi di controllo fu assegnato lo stesso tipo di test settimanale ma senza retroazione. Nessun test numerico fu assegnato nel corso delle otto settimane.

Nella decima e ultima sessione di test furono assegnati a tutt'e quattro i gruppi gli stessi test relativi al concetto ordinale, all'abilità nelle operazioni con i numeri naturali e al concetto cardinale che erano stati sottoposti loro nel corso del primo incontro e i risultati dei test finali furono analizzati statisticamente. L'analisi rivelò cinque scoperte principali (*si veda la figura a pagina 90*). Innanzitutto, i risultati ottenuti dai 60 bambini del primo gruppo sperimentale in relazione al concetto ordinale di numero furono nettamente superiori a quelli ottenuti dai 60 bambini del gruppo di controllo corrispondente, dimostrando che la nozione ordinale migliora nettamente in seguito a un addestramento fondato sulla semplice retroazione. In secondo luogo, anche i risultati ottenuti nei test relativi al concetto cardinale di numero dai bambini del secondo gruppo sperimentale furono superiori a quelli del relativo gruppo di controllo, dimostrando cosí che anche la nozione cardinale di numero migliora in seguito a un addestramento fondato sulla retroazione.

La terza scoperta fu che il miglioramento medio ottenuto nei test relativi alla concezione ordinale di numero fu molto maggiore di quello ottenuto mediamente in relazione alla concezione cardinale, indicando cosí che è piú difficile acquisire la nozione cardinale di quella ordinale. Le nostre ultime due scoperte riguardarono l'abilità numerica. Trovammo che i bambini appartenenti al gruppo sperimentale che era stato addestrato a test sul concetto ordinale di numero avevano ottenuto nelle operazioni con numeri risultati superiori ai bambini del relativo gruppo di controllo, mentre non erano significativamente superiori a quelli del relativo gruppo di controllo i risultati nelle operazioni con numeri ottenuti dai bambini del gruppo sperimentale che era stato addestrato con test relativi alla concezione cardinale di numero. Queste scoperte confermarono quanto era stato suggerito dai nostri studi normativi: ossia che la concezione dei numeri naturali deriva da una preliminare comprensione della concezione ordinale e non di quella cardinale.

È istruttivo considerare il problema del modo migliore in cui insegnare ai bambini le operazioni con i numeri alla luce di queste scoperte. Ci piaccia o no, l'abilità numerica è un requisito sociale nella civiltà moderna. Si potrebbe dire addirittura che l'abilità di operare con i numeri ha acquisito un significativo valore di sopravvivenza. Per esempio, alcune tra le decisioni piú importanti prese da una persona media nella sua vita implicano idee che hanno relazione con i numeri, come per esempio l'idea di denaro. È chiaro tuttavia che se non progettiamo l'istruzione matematica in modo che sia in accordo con i dati di fatto dello sviluppo mentale, correremo il rischio di rendere eccessivamente difficile per i bambini le prime fasi dell'istruzione matematica. Peggio ancora, correremo il rischio di scoraggiare un gran numero di quegli stessi bambini che stiamo cercando di istruire.

Nel XIX secolo e per gran parte del XX fu uso costante nelle nazioni occidentali cominciare l'insegnamento formale della matematica ai bambini con i numeri naturali. Fu questo l'uso deplorato da Russell nell'introduzione all'edizione riveduta dei *Principles of Mathematics* pubblicata nel 1937. Russell sostenne che tale uso derivava dallo screditato assunto dei pitagorici che i numeri naturali fossero entità non analizzabili che noi dobbiamo accettare semplicemente come date. Non c'è nulla di azzardato nel supporre che egli avrebbe preferito un piano di studi in cui l'introduzione dei numeri naturali fosse posticipata fino a quando il bambino avesse appreso i rudimenti della logica delle classi e, in particolare, la quantificazione di classi attraverso la corrispondenza dei loro elementi. L'assunto pitagorico era però ben trincerato nell'ambiente della scuola e le argomentazioni di Russell furono ignorate per circa un quarto di secolo.

Nell'ultimo decennio ha cominciato a imporsi il mutamento invocato da Russell. La maggior parte dei genitori, pur non rendendosi conto dei particolari, ricorda la parola d'ordine del periodo: « nuova matematica ». Nella « nuova matematica » era stata infine introdotta nell'insegnamento la teoria logicistica secondo cui i numeri naturali sono derivati dalla quantificazione di classi. Di fatto, sarebbe piú giusto definire la « nuova matematica » come l'approccio di tipo cardinale alle prime fasi dell'istruzione matematica; secondo quest'impostazione il bambino viene introdotto dapprima all'idea centrale della logica delle classi, e i numeri naturali vengono introdotti solo in seguito come semplici conseguenze del processo delle classi quantificanti.

Oggi, pur permanendo ancora sacche di resistenza, l'approccio di tipo cardinale ha nell'insegnamento della matematica una posizione di forza paragonabile a quella che aveva l'approccio pitagorico quando Russell si espresse contro di esso. La maggior parte dei piú importanti editori di libri di testo elementari hanno adottato o stanno adottando una qualche versione dell'approccio di tipo cardinale ed è diventata molto difficile una resistenza protratta al mutamento.

Tenendo conto delle implicazioni dei nostri esperimenti normativi e di trasferimento, appare pericoloso accontentarsi o del vecchio approccio pitagorico o del nuovo approccio di tipo cardinale. L'ordine in cui i vari concetti emergono nel pensiero del bambino è, come abbiamo visto: prima il concetto ordinale, poi la capacità di operare con i numeri e infine il concetto cardinale. L'addestramento nella concezione ordinale contribuisce inoltre a migliorare nei bambini la capacità di operare con i numeri, mentre lo stesso non si può dire per la concezione cardinale. Appare pertanto inevitabile la conclusione della necessità di una nuova « nuova matematica » che metta l'accento su nozioni ordinali.

L'asimmetria del cervello umano

Gli emisferi cerebrali, per quanto fisicamente simili, hanno funzioni diverse: il destro è specializzato per analizzare le informazioni sull'ambiente, il sinistro per atti motori specializzati, parola compresa

di Doreen Kimura

Nella maggior parte degli animali la struttura del sistema nervoso è essenzialmente simmetrica. Nei mammiferi la simmetria è resa piú evidente dalla prominenza della parte superiore del cervello: gli emisferi cerebrali. Tuttavia nell'uomo i due emisferi cerebrali differiscono notevolmente per le loro funzioni. È cosa ben nota che l'emisfero sinistro ha un ruolo dominante nel linguaggio (si veda l'articolo *Il linguaggio e il cervello* di Norman Geschwind, in « Le Scienze », n. 47, luglio 1972). L'emisfero destro ha anche delle funzioni specializzate, ma a questo riguardo fino a poco tempo fa avevamo poche informazioni perché gli studi erano concentrati sui disturbi del linguaggio. Ora si sa anche che l'emisfero destro ha nell'uomo un ruolo dominante nella percezione dell'ambiente.

Per piú di un secolo le fonti principali di conoscenza sulla divisione del lavoro tra i due emisferi cerebrali dell'uomo sono state le disfunzioni cerebrali provocate da incidenti, operazioni chirurgiche o malattie. Sebbene gli studi sulle alterazioni dell'intelligenza in pazienti con vari tipi di lesioni cerebrali abbiano fornito informazioni di un certo interesse, questi studi hanno lo svantaggio che il danno può aver colpito non solo gli specifici sistemi funzionali, ma anche le loro interazioni. Negli ultimi anni mi sono occupata di sviluppare metodi per lo studio delle asimmetrie delle funzioni emisferiche in persone normali.

Mi accorsi per la prima volta della possibilità che alcuni aspetti della funzione cerebrale potevano essere facilmente studiati in persone normali mentre facevo delle ricerche su pazienti all'Istituto neurologico di Montreal. Una delle prove alle quali venivano sottoposti i pazienti era una modificazione della tecnica di ascolto biauricolare ideata da Donald E. Broadbent del Centro di ricerca psicologica applicata del British Medical Research Council. La sua tecnica consiste nel far ascoltare simultaneamente un numero diverso a ciascun orecchio. Durante una prova, generalmente, vengono fatte ascoltare successivamente tre coppie di numeri e al soggetto viene chiesto di riferire tutti i numeri che ha udito. I pazienti con danni alla regione temporale sinistra del cervello riferiscono in modo corretto meno numeri dei pazienti con un danno alla regione temporale destra. Un risultato oscuro e del tutto inatteso fu che la maggior parte dei pazienti, indipendentemente da quale fosse la parte del cervello danneggiata, riferiva le parole che aveva udito con l'orecchio destro con maggior precisione di quelle che aveva udito con l'orecchio sinistro. Si vide che lo stesso avveniva per un gruppo di persone normali. In seguito a numerosi studi sulle soglie di tonalità pura si è provato che l'orecchio sinistro e il destro non differiscono per la loro capacità fondamentale di individuare i suoni, e cosí concludemmo che la superiorità di percezione delle parole dell'orecchio destro era in qualche modo in relazione con le connessioni dell'orecchio col cervello.

Una particolarità del sistema nervoso dell'uomo è che ogni emisfero cerebrale riceve informazioni essenzialmente dalla metà opposta del corpo. Il sistema visivo dell'uomo è fatto in modo tale che la visione alla destra di un punto fisso è mediata dalla metà sinistra del cervello e viceversa. Il sistema uditivo è in una certa misura meno incrociato poiché ogni metà del cervello riceve stimoli da entrambi gli orecchi, ma le connessioni incrociate sono tuttavia piú forti di quelle non incrociate. Il sistema tattile e quello motorio del cervello sono quasi completamente incrociati: le sensazioni e i movimenti della metà sinistra del corpo sono dipendenti essenzialmente dall'emisfero cerebrale destro e viceversa. I due emisferi stessi sono interconnessi mediante vie nervose. Queste vie, la piú grande delle quali è il corpo calloso, svolgono un ruolo importante nel coordinare le attività degli emisferi.

Poiché il sistema uditivo è essenzialmente un sistema incrociato, lo stimolo nervoso dell'orecchio destro all'emisfero cerebrale sinistro dovrebbe essere piú forte di quello dell'orecchio destro all'emisfero destro. E poiché l'emisfero sinistro contiene generalmente il sistema della percezione del linguaggio, è ragionevole pensare che i suoni del linguaggio offerti all'orecchio destro abbiano un accesso piú facile al sistema di percezione del linguaggio. Questa supposizione può essere provata direttamente osservando persone nelle quali le funzioni del linguaggio non si trovano nell'emisfero sinistro, ma nel destro. Si può stabilire da quale lato dipenda il linguaggio con la prova dell'amytal sodico, ideata da Juhn A. Wada. La prova implica l'iniezione di amytal sodico (un sedativo) in una sola delle due carotidi. La sostanza altera il funzionamento dell'emisfero cerebrale corrispondente a quel lato per pochi minuti, e se anche il linguaggio del soggetto viene alterato si deduce che il linguaggio dipende da quell'emisfero.

Si trovò che tredici pazienti che io esaminai all'Istituto neurologico di Montreal avevano un linguaggio dipendente dall'emisfero destro, invece che dal sinistro. Le prestazioni di questi pazienti nella prova uditiva biauricolare erano migliori per l'orecchio sinistro. Ciò sosteneva l'ipotesi che la superiorità dell'orecchio destro nei soggetti normali dipende da migliori connessioni tra quell'orecchio e l'emisfero sinistro (linguaggio) rispetto a ciò che avviene tra lo stesso orecchio e l'emisfero destro.

Trovammo ulteriori prove che le superiori prestazioni di un orecchio nelle prove uditive biauricolari riflettevano effettivamente una specializzazione funzionale degli emisferi. Brenda Milner dell'Istituto neurologico di Montreal trovò che mentre un danno del lobo temporale sinistro del cervello alterava la comprensione di materiale parlato, il danno del lobo temporale destro alterava la percezione di alcuni tipi di materiale uditivo, in particolare la discriminazione di qualità e caratteristiche di tonalità. Ho messo a punto una prova uditiva biauricolare in cui veniva utilizzata una cuffia per far ascoltare simultaneamente due melodie diverse ai due orecchi. Al soggetto veniva poi chiesto di scegliere le due melodie che aveva udito tra quattro melodie ognuna delle quali veniva fatta ascoltare singolarmente a entrambi gli orecchi. Poiché le melodie venivano elaborate essenzialmente dal lobo temporale destro, i soggetti erano in grado di scegliere la melodia offerta all'orecchio sinistro meglio di quella offerta all'orecchio destro.

I risultati erano particolarmente interessanti poiché aprivano la strada all'esplorazione delle caratteristiche dei processi verbali e non verbali del cervello per mezzo di tecniche relativamente semplici. Sebbene sia noto da piú di un secolo che l'emisfero sinistro è implicato nelle funzioni del linguaggio, ancora non abbiamo un'idea molto chiara di quali siano le caratteristiche di queste funzioni. Il modo tradizionale per distinguerle è di usare un termine come « simbolico » che comporta che le caratteristiche determinanti hanno a che fare con la capacità di far sí che un evento significhi qualcosa di diverso. Quando abbiamo usato la tecnica biauricolare dell'ascolto, abbiamo ottenuto una risposta piuttosto diversa. Si trovò che l'orecchio destro era superiore per sillabe senza significato e per suoni senza senso (come discorsi registrati trasmessi alla rovescia o una lingua straniera sconosciuta al soggetto). Anche Donald Shankweiler e Michael Studdert-Kennedy dei Laboratori Haskins di New Haven hanno applicato il metodo biauricolare al problema di definire le caratteristiche del linguaggio. Essi hanno trovato che non c'è alcuna superiorità dell'orecchio destro per quanto riguarda la percezione delle vocali isolate, ma che un tale effetto sussisteva per sillabe con consonanti e vocali. È difficile conciliare tutte queste scoperte con la nozione che il sistema di linguaggio dell'emisfero sinistro elabora essenzialmente materiale simbolico. Perché le vocali, che hanno valore simbolico, dovrebbero essere trattate allo stesso modo da entrambi gli emisferi, mentre suoni senza senso, come discorsi trasmessi alla rovescia, apparentemente vengono elaborati essenzialmente dall'emisfero sinistro?

Si è costretti a concludere che nel-

La prova di ascolto biauricolare consiste nel far ascoltare una melodia a un orecchio e un'altra, diversa, all'altro orecchio. Al soggetto viene poi chiesto di scegliere queste melodie in una serie di quattro melodie fatte ascoltare una alla volta a entrambi gli orecchi. L'ascolto avviene mediante auricolari connessi con un registratore a doppio canale. Altre prove implicano l'ascolto di numeri, parole, sillabe, vocali, discorsi trasmessi alla rovescia e suoni vocali come risate e pianti.

la percezione uditiva, l'emisfero sinistro è specializzato nella percezione di certi tipi di suoni generati dalle corde vocali e dal tratto vocale dell'uomo. Tagliando un nastro con la registrazione di un discorso normale in piccoli pezzi, Laurian King e io trovammo che la durata piú breve che dava luogo a una superiorità dell'orecchio destro era di circa 200 millisecondi, che è circa la durata media di una sillaba: una consonante e una vocale. Tale misura di unità sembra necessaria, sebbene non sempre sufficiente, per elaborazioni asimmetriche ed è in favore della nozione che la sillaba è una unità basilare del linguaggio.

Abbiamo studiato inoltre la percezione biauricolare di diversi suoni vocalici non appartenenti al linguaggio, come tosse, risate e pianto. Invece di trovare una superiorità dell'orecchio destro per questi suoni, abbiamo riscontrato una superiorità dell'orecchio sinistro. Questo risultato fa pensare che questi suoni siano elaborati principalmente dall'emisfero destro, proprio come avviene per i suoni musicali. Anche i motivi canticchiati da un'altra persona venivano meglio identificati se ascoltati con l'orecchio sinistro. Perciò, se l'emisfero sinistro distingue i suoni per le loro caratteristiche articolatorie, queste caratteristiche devono essere piuttosto specifiche.

Il linguaggio va inteso generalmente, con alcune riserve, come un sistema essenzialmente uditivo-vocale. Per quanto la comprensione di una parola scritta sia conseguenza di una esperienza precedente con l'equivalente parlato, ci si può aspettare che il sistema del linguaggio prenda parte all'elaborazione di materiale stampato e scritto; infatti, quando si fanno degli esami approfonditi della capacità di lettura, i disturbi della parola sono quasi sempre accompagnati almeno da un leggero disturbo nella lettura. Cominciammo a cercare se esistessero oppure no delle asimmetrie nella percezione visiva analoghe all'effetto orecchio destro-orecchio sinistro che avevamo riscontrato nelle modalità uditive. Alcuni studi precedenti avevano infatti dimostrato che parole e lettere venivano riportate con piú precisione dalla metà destra del campo visivo, piuttosto che dalla metà sinistra, ma questi effetti erano spiegati come derivati da abitudini di lettura. Al fine di mettere in relazione specificatamente simili effetti con l'asimmetria della funzione cerebrale, dovevamo trovare delle prove che interessassero anche le funzioni visive dell'emisfero destro.

Sebbene anche il sistema visivo sia incrociato, le sue connessioni sono diverse da quelle del sistema uditivo. Le connessioni non uniscono ciascun occhio alla metà opposta del cervello, ma ogni metà del campo visivo alla corteccia visiva del lato opposto (*si veda la figura nella pagina a fronte*). La percezione visiva a sinistra del punto di fissazione è ricevuta dalla metà destra di ogni retina e le vie nervose provenienti dalla parte destra di entrambe le retine vanno alla corteccia visiva dell'emisfero destro. Ovviamente, le fibre provenienti dalla metà destra della retina dell'occhio sinistro devono attraversare la linea mediana del cervello per portarsi all'emisfero destro, ma le fibre provenienti dalla metà destra della retina dell'occhio destro non devono incrociarsi.

Una differenza importante tra sistema uditivo e visivo è che, quando la testa è immobile, le orecchie sono ferme, ma gli occhi no. Poiché gli occhi sono in continuo movimento, in condizioni visive normali una immagine non può presentarsi in un solo campo visivo. Per superare questa difficoltà gli stimoli devono essere presentati con grande rapidità in un certo periodo quando il punto di fissazione è noto. Questo lo si può fare con lo strumento noto come tachiscopio (*si veda la figura a pagina 98*). Il soggetto guarda dentro l'apparecchio e gli viene richiesto di fissare un punto determinato. Mentre egli lo sta fissando, viene presentato molto rapidamente uno stimolo visivo sia a sinistra che a destra di quel punto. Prima che il soggetto possa spostare il suo punto di fissazione per guardare lo stimolo, questo è scomparso. Ne risulta che solo un lato di ciascuna delle retine e solo un emisfero cerebrale sono direttamente stimolati.

Nella maggior parte dei nostri esperimenti col tachiscopio, non abbiamo fatto uso di competizione tra due stimoli, come abbiamo fatto nelle prove uditive biauricolari. Infatti i nostri dati erano meno ambigui quando veniva stimolato un solo campo visivo. Nelle persone normali, poi, parole e lettere vengono riportate piú accuratamente dal campo visivo destro, piuttosto che dal sinistro. Questo dato è in accordo con la superiorità dell'orecchio destro nel riconoscimento di suoni parlati. Vale a dire che il riconoscimento di materiale visivo relativo al linguaggio è anche piú accurato quando questo materiale stimola inizialmente l'emisfero sinistro.

Le vie uditive che vanno dagli orecchi alle aree riceventi uditive del cervello negli emisferi destro e sinistro sono parzialmente incrociate. Sebbene ogni emisfero possa ricevere stimoli da entrambi gli orecchi, le connessioni neurali tra un orecchio e l'emisfero del lato opposto sono piú forti di quelle con l'emisfero dello stesso lato. Quando gli stimoli ipsilaterali (dello stesso lato) e controlaterali (del lato opposto) sono in competizione, si pensa che lo stimolo controlaterale inibisca o copra i segnali ipsilaterali.

Gran parte del nostro lavoro sulla percezione visiva è stato rivolto a scoprire alcune delle funzioni specializzate dell'emisfero destro. Era noto da molto tempo che un danno alla parte posteriore destra del cervello (la regione occipito-parietale) aveva come risultato l'alterazione di attività complesse come il disegno, il senso dell'orientamento e il costruire qualcosa in base a uno schema o a una figura. Nei nostri studi su soggetti normali abbiamo trovato prove che l'emisfero destro è anche essenziale per alcuni processi visivi molto importanti. Per esempio abbiamo trovato che nel tipo piú semplice di prova sul senso dello spazio – la localizzazione di un punto in un'area bidimensionale – è dominante l'emisfero destro. Saggiammo questa capacità presentando dei punti, uno alla volta, o nel campo visivo sinistro o in quello destro per un centesimo di secondo. Il punto veniva presentato in varie posizioni dentro un cerchio disegnato su un cartoncino liscio bianco. Il soggetto identificava quindi la posizione del punto su un cartoncino simile fuori del tachiscopio. I risultati per la corretta identificazione dei punti erano migliori per i punti presentati nel campo sinistro rispetto a quelli presentati nel campo destro. Inoltre, lo stabilire il numero dei punti e le forme geometriche era piú accurato per il campo sinistro.

Questa capacità non è dovuta solamente a una piú elevata attenzione per gli stimoli nel lato sinistro, poiché la semplice identificazione dei punti non è piú accurata in un campo rispetto all'altro. La capacità di identificazione veniva saggiata facendo riferire al soggetto se un punto era presente o meno in ogni prova con un tempo di esposizione fisso, o stabilendo il tempo di esposizione richiesto per l'identificazione di un punto. In nessun caso c'è differenza tra l'accuratezza di identificazione del campo sinistro e di quello destro. Sembra invece che l'emisfero destro racchiuda componenti importanti di un sistema di coordinate spaziali che facilita la localizzazione di un punto nello spazio. Naturalmente gli stimoli che giungono a entrambe le cortecce visive devono in genere avere accesso a questo sistema, ma quando lo stimolo è deliberatamente limitatc a un solo campo visivo è possibile determinare se un emisfero ha un vantaggio funzionale.

Ci chiedemmo quindi se l'emisfero destro poteva anche essere importante

Le vie ottiche sono completamente incrociate, in modo tale che, quando gli occhi sono fissi su di un punto, tutto ciò che si trova a sinistra del punto dove si fissa lo sguardo eccita la corteccia visiva dell'emisfero destro e gli stimoli del campo visivo destro eccitano la corteccia visiva sinistra. Le cortecce visive sono in grado di comunicare tra loro attraverso il corpo calloso che è la principale connessione tra i due emisferi cerebrali.

per la percezione di profondità. La localizzazione tridimensionale di oggetti può essere mediata da diversi fattori. Di questi molti sono monoculari, vale a dire che interessano un occhio solo. I fattori comprendono le dimensioni relative delle immagini retiniche, l'oscuramento di un soggetto da parte di un altro, o la velocità relativa con cui due oggetti si muovono nel campo visivo. Un altro fattore importante per la profondità, la disparità binoculare, interessa entrambi gli occhi. La disparità binoculare si riferisce al fatto che, poiché i due occhi sono separati, ogni occhio riceve un'immagine retinica leggermente diversa. La disparità tra le due immagini può fornire informazioni sulla profondità di un oggetto poiché gli oggetti piú vicini hanno una disparità binoculare maggiore di quelli piú lontani.

Margaret Durnford e io iniziammo degli studi sulla percezione della profondità attaccando una tipica camera per la percezione della profondità alla parte posteriore del tachiscopio. La camera contiene, posta al centro, una bacchetta verticale fissa in linea con il punto di fissazione. Su ciascun lato della bacchetta centrale c'è una scanalatura sulla quale si può muovere un'altra bacchetta verticale. La bacchetta mobile viene vista da entrambi gli occhi solo per una frazione di secondo, e al soggetto viene chiesto se era piú vicina o piú lontana della bacchetta centrale. Quando la bacchetta mobile si trovava nel campo visivo sinistro, vale a dire quando l'informazione andava all'emisfero destro, le risposte erano piú accurate. Cosí anche l'informazione spaziale sulla terza dimensione è elaborata con maggiore precisione dall'emisfero destro rispetto al sinistro.

Quando la bacchetta mobile era vista solo con l'occhio destro non c'era alcuna differenza nella precisione tra il campo sinistro e il destro. Questo risultato suggerí che l'informazione binoculare era elaborata principalmente dall'emisfero destro e che l'informazione monoculare era elaborata da ciascun emisfero. Come ulteriore prova della specificità della simmetria emisferica presentammo degli stimoli visivi in cui l'unico fattore importante per la profondità era la disparità binoculare. Quando si vedono separatamente con i due occhi mediante uno stereoscopio due immagini leggermente diverse di una figura bidimensionale, si può simulare la disparità binoculare naturale. Il soggetto riferisce di vedere le due immagini come fuse insieme e in possesso delle caratteristiche di profondità che le due immagini singole non hanno.

Gli stereogrammi con punti disposti alla rinfusa, ideati da Bela Julesz dei Laboratori Bell sono molto adatti a questo scopo, poiché ogni immagine monoculare appare come una disposizione di punti alla rinfusa. Quando le due immagini sono viste stereoscopicamente, appare, davanti o dietro alla figura con i punti, l'immagine di un quadrato o di un triangolo. Venne presentata una serie di stereogrammi simili con diverse forme, mediante un tachiscopio, all'occhio destro oppure al sinistro, e fu chiesto al soggetto di identificare l'immagine che risultava. Per esempio, in una prova interessante il campo sinistro, una immagine di punti alla rinfusa viene presentata nel campo sinistro dell'occhio sinistro e l'altro stereogramma nel campo sinistro dell'occhio destro. Le due figure presentate in

Il tachiscopio a due campi è utilizzato per lo studio della percezione visiva. Quando il campo di fissazione è illuminato, un osservatore vede il campo riflesso nello specchio semiargentato. Gli viene chiesto di fissare un punto al centro del campo. Poi la luce del campo di fissazione viene spenta e si accende contemporaneamente per pochi millisecondi la luce del campo di esposizione. L'immagine del campo di esposizione passa attraverso lo specchio semiargentato ed è vista per un momento dall'osservatore. Alla fine dell'esposizione, la luce del campo di fissazione si accende e quella del campo di esposizione si spegne. Ponendo l'immagine del campo di esposizione nel campo visivo sinistro, o destro a propria scelta, si possono stimolare selettivamente le due cortecce visive.

Il tachiscopio a tre campi viene utilizzato negli studi sulla percezione della profondità. Il soggetto fissa inizialmente un punto che vede riflesso da uno specchio parzialmente argentato. I due campi di esposizione diventano visibili quando la luce del campo di fissazione si spegne e le luci dei campi di esposizione si accendono per un istante. I filtri polarizzatori sono messi in modo tale che ciascun occhio riceve un'immagine diversa. Utilizzando una tecnica messa a punto da Bela Julesz dei Laboratori Bell, l'autrice mostrava un insieme di punti alla rinfusa alla emiretina destra dell'occhio destro e un insieme leggermente diverso alla emiretina destra dell'occhio sinistro. In altre prove l'insieme di punti era mostrato alle emiretine sinistre. Quando ogni insieme è visto singolarmente, non si notano né forme né profondità, ma quando i due insiemi di punti vengono in qualche modo fusi nel sistema visivo, il soggetto vede un triangolo fluttuante davanti all'insieme di punti. La maggior parte delle persone identifica meglio la figura stereoscopica quando le immagini sono proiettate nel campo visivo sinistro rispetto a quando sono proiettate nel destro, indicando che l'emisfero destro è in grado di elaborare meglio l'informazione di profondità.

questo modo sono inizialmente elaborate nell'emisfero cerebrale destro. Una presentazione nel campo destro, d'altro canto, porta a una elaborazione stereoscopica nell'emisfero sinistro.

Abbiamo anticipato alcune delle difficoltà che si hanno per ottenere la fusione binoculare a causa della breve esposizione (100 millisecondi) che dovevamo usare, ma, con una certa sorpresa per noi, la maggior parte dei soggetti che osservammo era in grado di svolgere la prova. La identificazione di stimoli stereoscopici era nettamente migliore quando gli stimoli venivano presentati nel campo visivo sinistro, vale a dire quando le immagini distinte andavano all'emisfero destro. Può darsi, quindi, che l'elaborazione delle informazioni di profondità nell'emisfero destro, sia connessa in modo piuttosto specifico con l'utilizzazione di tali fattori binoculari.

Risulta chiaro che l'emisfero destro funziona meglio dell'emisfero sinistro nell'analisi delle informazioni riguardanti la localizzazione spaziale di oggetti. La probabilità che si avesse a che fare con funzioni fondamentali per lo orientamento spaziale ci incoraggiò a studiare la specializzazione emisferica per un altro fondamentale processo visivo: la percezione dell'inclinazione di una retta. Venivano presentate delle linee molto corte, una alla volta, nel campo visivo sinistro o nel destro. Le linee variavano di inclinazione da 15 a 165 gradi, con intervalli di 15 gradi. Dopo che il soggetto aveva visto la linea, gli veniva chiesto di individuarla scegliendola in un insieme di linee inclinate tracciate su un un foglio di carta. C'era una piccola, ma significativa superiorità nell'identificazione dell'inclinazione per il campo visivo sinistro, suggerendo ancora una volta che questa informazione è elaborata nel cervello in modo asimmetrico. David H. Hubel e Torsten N. Wiesel della Harvard Medical School hanno proposto che l'elaborazione iniziale dell'orientamento di una linea si svolga nella corteccia visiva. Il fatto che noi abbiamo trovato una asimmetria emisferica per questa funzione relativa alla percezione, apre la possibilità che ci possa essere un certa asimmetria funzionale tra gli emisferi in processi fondamentali cosí come in processi associativi piú complessi.

Non sappiamo con precisione quale sistema neurale del cervello stiamo saggiando con le nostre tecniche sulla percezione. Le prove che abbiamo suggeriscono che con gli esami al tachiscopio tocchiamo le funzioni di regioni vicine alla corteccia striata, la piú importante via visiva per gli emisferi, piuttosto che regioni piú lontane come i lobi temporali che anche hanno delle funzioni visive. Con le prove uditive biauricolari, d'altra parte, tocchiamo probabilmente le funzioni dell'area del lobo temporale.

Un altro punto interessante, in relazione con quanto detto, è che non abbiamo trovato una qualsiasi superiorità del campo visivo sinistro per la percezione della forma, sebbene sia stato tentato con parecchi esperimenti. Sappiamo che un danno al lobo temporale destro altera la percezione di disegni privi di significato, suggerendo che alcune porzioni dell'emisfero destro possano avere una importanza critica per la percezione della forma. Apparentemente i sistemi neurali interessati alla elaborazione spaziale sono relativamente indipendenti da quelli interessati alla percezione della forma.

Sebbene noi ci siamo interessati in particolar modo alla percezione visiva e uditiva, sembra ci sia una analoga asimmetria nella percezione tattile. Beata Hermelin e Neil O'Connor del Centro di psicologia del Medical Research Council hanno riferito che la percezione tattile di disposizioni di punti Braille da parte di persone cieche è piú rapida con la mano sinistra che con la destra. Nel nostro laboratorio Diana Ingram ha trovato che, quando un braccio è utilizzato per localizzare un punto fuori dalla vista sotto una tavola sulla quale è indicata la posizione del punto, il braccio sinistro agisce con maggior precisione del destro.

Da questa e da altre osservazioni si può concludere che la parte posteriore dell'emisfero destro è implicata nell'analisi diretta delle informazioni riguardanti l'ambiente esterno. L'area occipito-parietale è particolarmente importante per i tipi di comportamento che dipendono da relazioni spaziali, mentre la regione temporale prende parte all'elaborazione di stimoli non spaziali come composizioni melodiche e disegni privi di significato. Un importante processo successivo nell'analisi dello stimolo relativo alla percezione è l'aggiunta di un contrassegno verbale. Sappiamo che la trasformazione verbale dell'informazione interessa l'emisfero sinistro, ma abbiamo ancora molto da imparare sul trasferimento della elaborazione dall'emisfero destro al sinistro e viceversa. Per esempio, possiamo dimostrare la superiorità del campo visivo sinistro nella percezione spaziale, anche se il soggetto ci dà in effetti una risposta verbale che è controllata dal suo emisfero sinistro. Spieghiamo questo risultato dicendo che l'analisi principale è compiuta dall'emisfero destro e la risposta verbale è secondaria. In altre situazioni, tuttavia, le modalità di risposta (manuale o vocale) possono influenzare il campo che domina.

Abbiamo visto che l'emisfero sinistro è essenziale per la produzione e la percezione di alcuni suoni emessi dal sistema umano del linguaggio. Recentemente con i miei colleghi ho ottenuto prove, con pazienti colpiti da apoplessia, che anche l'emisfero sinistro può essere essenziale per alcuni tipi di movimenti della mano. Abbiamo trovato che pazienti con danni alla metà sinistra del cervello avevano difficoltà a eseguire una serie di movimenti della mano senza tener conto del significato dei movimenti stessi. Inoltre, ci sono dei resoconti nella letteratura clinica, riguardanti sordomuti, che utilizzavano i movimenti della mano come mezzo di comunicazione, i quali, dopo aver subito un danno all'emisfero sinistro, mostravano alterazioni di questi movimenti analoghe ai disturbi del linguaggio. Che l'emisfero sinistro eserciti un particolare controllo su certi aspetti del comportamento manuale è ulteriormente suggerito dal fatto che la maggior parte delle persone usa la mano destra per molte azioni che richiedono abilità. Sebbene la relazione tra lateralizzazione del linguaggio e preferenza nell'uso della mano non sia perfetta, l'elevata frequenza sia di controllo del linguaggio a opera dell'emisfero sinistro che di preferenza nell'uso della mano destra non è probabilmente una coincidenza.

Abbiamo trovato un ulteriore sostegno per la relazione tra linguaggio e alcune attività manuali osservando i movimenti delle mani di persone normali mentre esse parlavano. Come tutti sanno, il discorso è spesso accompagnato da gesti, in cui le mani si muovono liberamente nello spazio senza toccare nulla. Questi movimenti sono molto rari in attività vocali non appartenenti al linguaggio, come quando si canticchia a bocca chiusa. Tuttavia, in questo ultimo caso, e quando si parla, ci possono essere altri tipi di attività manuale, come toccarsi il corpo, strofinarsi il naso o grattarsi. Ugualmente interessante è il fatto che i gesti che si fanno mentre si parla sono fatti essenzialmente dalla mano opposta all'emisfero che controlla il linguaggio (determinato mediante il sistema verbale biauricolare). Se il linguaggio è controllato dall'emisfero sinistro, come è nella maggior parte delle persone, la mano destra compie il maggior numero di gesti; se è l'emisfero destro che lo controlla, allora è la mano sinistra ad agitarsi di piú. È un fatto curioso che questa asimmetria sia ristretta solo ai

I gesti durante un discorso sono fatti essenzialmente con la mano che si trova dalla parte opposta rispetto all'emisfero cerebrale che controlla il linguaggio. Nella maggior parte delle persone il linguaggio è controllato dall'emisfero sinistro e la mano destra di queste persone compie, durante il discorso, un maggior numero di movimenti liberi rispetto alla mano sinistra, come si vede da queste immagini prese da una registrazione televisiva fatta nel corso di un esperimento. L'uso asimmetrico delle mani durante il discorso è stato rilevato solo quando le mani si muovono liberamente nello spazio senza toccare il corpo.

Nelle persone normali, che utilizzano la mano destra, si ritrovano nelle modalità uditive, visive e manuali, le asimmetrie funzionali degli emisferi cerebrali. I punteggi relativi ai lati sinistro e destro raccolti nel corso delle prove sono stati convertiti in rapporti per un più facile confronto. Il rapporto per la dominanza dell'emisfero sinistro nel caso della percezione di parole parlate è 1,88 : 1, mentre il rapporto per la dominanza dell'emisfero destro nel caso di melodie è di 1,19 : 1. Questi rapporti non sono però valori fissi poiché variano col tipo di stimolo, col tipo di risposta richiesta e con la difficoltà della prova.

movimenti liberi; non sembra riguardare i movimenti che interessano il corpo.

Nel nostro lavoro sulla asimmetria cerebrale in persone normali abbiamo talvolta incontrato delle differenze in rapporto al sesso. Nelle prove interessanti l'emisfero destro, i maschi tendono ad avere una superiorità maggiore per il campo visivo sinistro nella localizzazione e nel conteggio di punti rispetto alle femmine. Sappiamo anche che i maschi sono superiori alle femmine in certe prove spazio-visive. Può essere che la specializzazione dell'emisfero destro sia più pronunciata nei maschi che nelle femmine e che questa specializzazione talvolta sia vantaggiosa. Di recente, nel mio laboratorio all'Università dell'Ontario occidentale, Jeannette McGlone e Wilda Davidson hanno trovato delle prove a favore di questa idea. Esse hanno eseguito una tipica prova di percezione spaziale in cui il soggetto deve identificare un disegno dopo che è stato ruotato. Esse trovarono, come è solito in questa prova, che i maschi generalmente si comportavano meglio delle femmine. Le femmine che avevano dei risultati particolarmente scadenti nella prova erano quelle che avevano una certa specializzazione dell'emisfero sinistro per funzioni spaziali (come suggerito dalla prova di conteggio dei punti al tachiscopio). In genere queste funzioni sono controllate dall'emisfero destro.

Al contrario, le femmine tendono ad avere una maggior fluidità verbale rispetto ai maschi. Tuttavia, non ci sono prove che le femmine adulte siano più asimmetriche dei maschi nella lateralizzazione del linguaggio. Gli studi biauricolari, però, fanno pensare che la lateralizzazione del linguaggio possa svilupparsi prima nelle ragazze che nei ragazzi. Sembra che, per alcune funzioni intellettuali, il cervello maschile e quello femminile siano organizzati diversamente. La maggior parte dell'evoluzione dell'uomo si deve essere svolta in condizioni in cui per i componenti maschili cacciatori era di estrema importanza avere delle informazioni sull'ambiente circostante vicino e lontano. Per le femmine, che presumibilmente rimanevano vicine alla casa con altri componenti non cacciatori del gruppo, dei processi di selezione simili possono non aver operato. Sarà interessante scoprire se le differenze sessuali nelle asimmetrie concernenti o non concernenti il linguaggio che abbiamo trovato con le nostre tecniche, relativamente semplici, siano vere oppure no anche per altre culture.

RAPPORTO TRA I PUNTEGGI DELLE PROVE

		DOMINANZA DELL'EMISFERO SINISTRO	DOMINANZA DELL'EMISFERO DESTRO
PROVE UDITIVE	PAROLE	1,88	
	SILLABE PRIVE DI SIGNIFICATO	1,73	
	DISCORSI ALLA ROVESCIA	1,66	
	MELODIE		1,19
	SUONI NON APPARTENENTI AL LINGUAGGIO (UMANI)		1,08
PROVE VISIVE	LETTERE	1,23	
	PAROLE	1,47	
	LOCALIZZAZIONE BIDIMENSIONALE DEI PUNTI		1,18
	ENUMERAZIONE DI PUNTI E DI FORME		1,20
	APPAIAMENTO DI LINEE INCLINATE		1,05
	PERCEZIONE STEREOSCOPICA DI PROFONDITÀ		1,28
PROVE MANUALI	MOVIMENTI DI ABILITÀ	1,13	
	MOVIMENTI LIBERI DURANTE IL DISCORSO	3,10	
	DISPOSIZIONI TATTILI DI PUNTI (BRAILLE)		NESSUN DATO
	LOCALIZZAZIONE NON VISIVA		1,12

Nelle persone normali, che utilizzano la mano destra, si ritrovano nelle modalità uditive, visive e manuali, le asimmetrie funzionali degli emisferi cerebrali. I punteggi relativi ai lati sinistro e destro raccolti nel corso delle prove sono stati convertiti in rapporti per un più facile confronto. Il rapporto per la dominanza dell'emisfero sinistro nel caso della percezione di parole parlate è 1,88 : 1, mentre il rapporto per la dominanza dell'emisfero destro nel caso di melodie è di 1,19 : 1. Questi rapporti non sono però valori fissi poiché variano col tipo di stimolo, col tipo di risposta richiesta e con la difficoltà della prova.

movimenti liberi; non sembra riguardare i movimenti che interessano il corpo.

Nel nostro lavoro sulla asimmetria cerebrale in persone normali abbiamo talvolta incontrato delle differenze in rapporto al sesso. Nelle prove interessanti l'emisfero destro, i maschi tendono ad avere una superiorità maggiore per il campo visivo sinistro nella localizzazione e nel conteggio di punti rispetto alle femmine. Sappiamo anche che i maschi sono superiori alle femmine in certe prove spazio-visive. Può essere che la specializzazione dell'emisfero destro sia più pronunciata nei maschi che nelle femmine e che questa specializzazione talvolta sia vantaggiosa. Di recente, nel mio laboratorio all'Università dell'Ontario occidentale, Jeannette McGlone e Wilda Davidson hanno trovato delle prove a favore di questa idea. Esse hanno eseguito una tipica prova di percezione spaziale in cui il soggetto deve identificare un disegno dopo che è stato ruotato. Esse trovarono, come è solito in questa prova, che i maschi generalmente si comportavano meglio delle femmine. Le femmine che avevano dei risultati particolarmente scadenti nella prova erano quelle che avevano una certa specializzazione dell'emisfero sinistro per funzioni spaziali (come suggerito dalla prova di conteggio dei punti al tachiscopio). In genere queste funzioni sono controllate dall'emisfero destro.

Al contrario, le femmine tendono ad avere una maggior fluidità verbale rispetto ai maschi. Tuttavia, non ci sono prove che le femmine adulte siano più asimmetriche dei maschi nella lateralizzazione del linguaggio. Gli studi biauricolari, però, fanno pensare che la lateralizzazione del linguaggio possa svilupparsi prima nelle ragazze che nei ragazzi. Sembra che, per alcune funzioni intellettuali, il cervello maschile e quello femminile siano organizzati diversamente. La maggior parte dell'evoluzione dell'uomo si deve essere svolta in condizioni in cui per i componenti maschili cacciatori era di estrema importanza avere delle informazioni sull'ambiente circostante vicino e lontano. Per le femmine, che presumibilmente rimanevano vicine alla casa con altri componenti non cacciatori del gruppo, dei processi di selezione simili possono non aver operato. Sarà interessante scoprire se le differenze sessuali nelle asimmetrie concernenti o non concernenti il linguaggio che abbiamo trovato con le nostre tecniche, relativamente semplici, siano vere oppure no anche per altre culture.

GIOCHI MATEMATICI

di Martin Gardner

Nuovi rompicapo tratti dal gioco dell'alma, il nobile antenato della dama cinese

Recentemente sono venute in luce due famiglie di rompicapo fondati su un gioco analogo alla dama che è stato trascurato per molto tempo. Ciascuna di queste famiglie ci offre una serie di problemi non risolti e l'opportunità di escogitare ingegnose dimostrazioni del fatto che alcune soluzioni sono impossibili. Questi problemi sono tratti da *Dialogo sui rompicapo*, una splendida raccolta di problemi insoliti edita nel 1971 in Giappone a cura di Kobon Fujimura e Michio Matsuda. (Il libro non è purtroppo disponibile nelle lingue occidentali.)

I due giochi per scacchiera che presentiamo sono derivati da un problema creato da Matsuda che sfrutta le semplici regole di un gioco popolare verso la fine del XIX secolo in Gran Bretagna e chiamato « alma » dalla parola greca che significa « salto ».

La scacchiera tradizionale per questo gioco ha 16 caselle per ogni lato (*si veda la figura in basso*). Se i giocatori sono due, ognuno comincia collocando le sue 19 pedine in una sezione della scacchiera chiamata « campo ». Ci sono due campi, uno all'angolo superiore sinistro della scacchiera e l'altro all'angolo inferiore destro. Le pedine sono identiche, tranne naturalmente la differenza di colore per i due avversari. Obiettivo del gioco è l'occupazione del campo dell'avversario, e il giocatore che per primo avrà portato tutte le sue pedine nel campo nemico sarà il vincitore. Sono consentiti due tipi di mosse: 1) Un « passo ». Il passo consiste nello spostamento da una casella a una delle otto caselle contigue: ciascuna pedina può muovere praticamente come il re negli scacchi. 2) Un « salto ». Questa mossa consente di scavalcare un'altra pedina, come nella dama, con la differenza che il salto può esser fatto in ogni direzione, ossia non solo in diagonale, ma anche ortogonalmente. Il pezzo saltato non viene mangiato ma rimane in gioco.

Una serie di salti collegati vale come una sola mossa. Non è obbligatorio saltare. Un giocatore può continuare una serie di salti finché gli è possibile, oppure può fermarsi quando gli pare. Non ha nessuna importanza il colore del pezzo che viene saltato; una serie di salti può superare pedine amiche e nemiche. In una stessa mossa non si possono però combinare passi e salti. I giocatori muovono alternativamente, spostando una pedina per volta.

Il gioco dell'alma può essere giocato anche da quattro persone: in questo caso ciascun contendente ha 13 pedine. I campi si trovano ai quattro angoli della scacchiera, dietro i limiti indicati nella figura dalle linee a trattini. Quando i contendenti sono quattro, ognuno può giocare per sé, cercando ciascuno di raggiungere il campo diagonalmente opposto al proprio; oppure possono formarsi coppie di giocatori opposti (o adiacenti) che si aiutino l'un l'altro; in questo caso vincerà la coppia che farà pervenire per prima negli opposti campi tutte e 26 le proprie pedine. La strategia del gioco è però così complessa che il gioco riesce molto più bello quando giocano due sole persone.

Tra i molti giochi successivi fondati sull'alma, i due più popolari negli Stati Uniti sono stati il *camelot* e la dama cinese, apparsi entrambi sul mercato americano fra il 1930 e il 1940. Il *camelot* era una ripresa (con differenze di poco conto) a opera dei fratelli Parker di un gioco della fine del XIX secolo. La dama cinese, che non ha nessun rapporto con la Cina, si gioca su una scacchiera dalle caselle esagonali che ha di solito la forma di una stella a sei punte. La struttura esagonale consente passi e salti in sei sole direzioni. Una versione francese dell'alma, nota

La scacchiera per il gioco dell'alma.

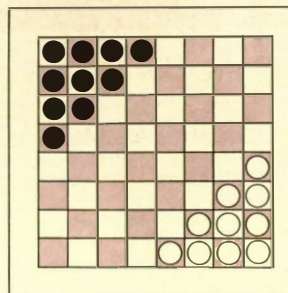
come *cavalletta*, può essere giocata su una normale scacchiera (*si veda la figura in alto nella pagina a fronte*).

Nei giochi del tipo dell'alma, per impedire a un giocatore ostinato di imporre un pari tenendo una pedina permanentemente nel proprio campo, è opportuno aggiungere altre norme. Sidney Sackson, di New York, inventore e collezionista di giochi, fece la seguente proposta. Se una pedina può uscire dal proprio campo saltando una pedina avversaria o con una serie di salti che comincino saltando una pedina avversaria, è obbligata a farlo, anche se, una volta fuori del campo, può interrompere dove desidera la sua serie di salti. Dopo che una pedina è uscita dal proprio campo, non può tornare a fermarsi in esso, ma può attraversarlo.

Il problema dell'alma escogitato da Matsuda per la scacchiera giapponese, che ha nove caselle per ogni lato, comincia con nove pedine disposte in un quadrato di tre pedine di lato all'angolo sinistro in basso della scacchiera. Qual è il numero minimo di mosse, si chiese Matsuda, per trasferire le nove pedine, nella stessa formazione, nell'angolo destro superiore? La soluzione da lui data fu di 17 mosse, ma il totale fu ridotto a 16 mosse (*si veda la serie di figure qui sotto*) da H. Ajisawa e da T. Maruyama. Si ritiene che la soluzione in 16 mosse sia quella che utilizza il minimo numero possibile di mosse.

Quando vidi questa elegante soluzione, subito cominciai ad affrontare lo stesso problema sulla scacchiera occidentale con otto caselle per ogni lato e su scacchiere più piccole, con sette e sei caselle per lato. Usando la tecnica di stabilire innanzitutto una scala in diagonale – una strategia fondamentale, sia detto per inciso, in tutti i giochi del tipo dell'alma – i risultati migliori che potei ottenere furono 15 per la scacchiera standard, 13 per quella con 7 caselle per lato e 12 per quella con 6. Non sono stato in grado di dimostrare per nessuna di queste soluzioni se si tratta del numero minimo di mosse possibile. Non è difficile far vedere che sono necessarie almeno 12 mosse per la scacchiera con 8 caselle per lato, 10 per quella di 7 caselle e 11 per quella di 6.

La soluzione del problema di Matsuda sulla scacchiera giapponese.

Successivamente sperimentai il problema analogo di trasferire le nove pedine, sulle stesse tre scacchiere, dall'angolo inferiore sinistro all'angolo inferiore destro, invece che a quello superiore destro, diagonalmente opposto. Nel caso della scacchiera con lato di 6 caselle esistono molte soluzioni in nove mosse: una di queste è illustrata nella figura in basso. Nove è ovviamente il numero di mosse minimo perché ogni pedina deve muovere almeno una volta. (Almeno una pedina, nel passaggio all'altro campo, deve saltare nella

Il gioco della cavalletta.

quarta riga e poi tornare indietro: perciò la soluzione in nove mosse non è possibile su una scacchiera rettangolare di sei caselle per tre.) Sulla scacchiera di 7 caselle di lato, il numero minimo di mosse è 10, fatto facilmente comprensibile se si pensa che il primo pezzo deve muovere almeno due volte per raggiungere il campo avverso.

Tredici mosse consentono di risolve-

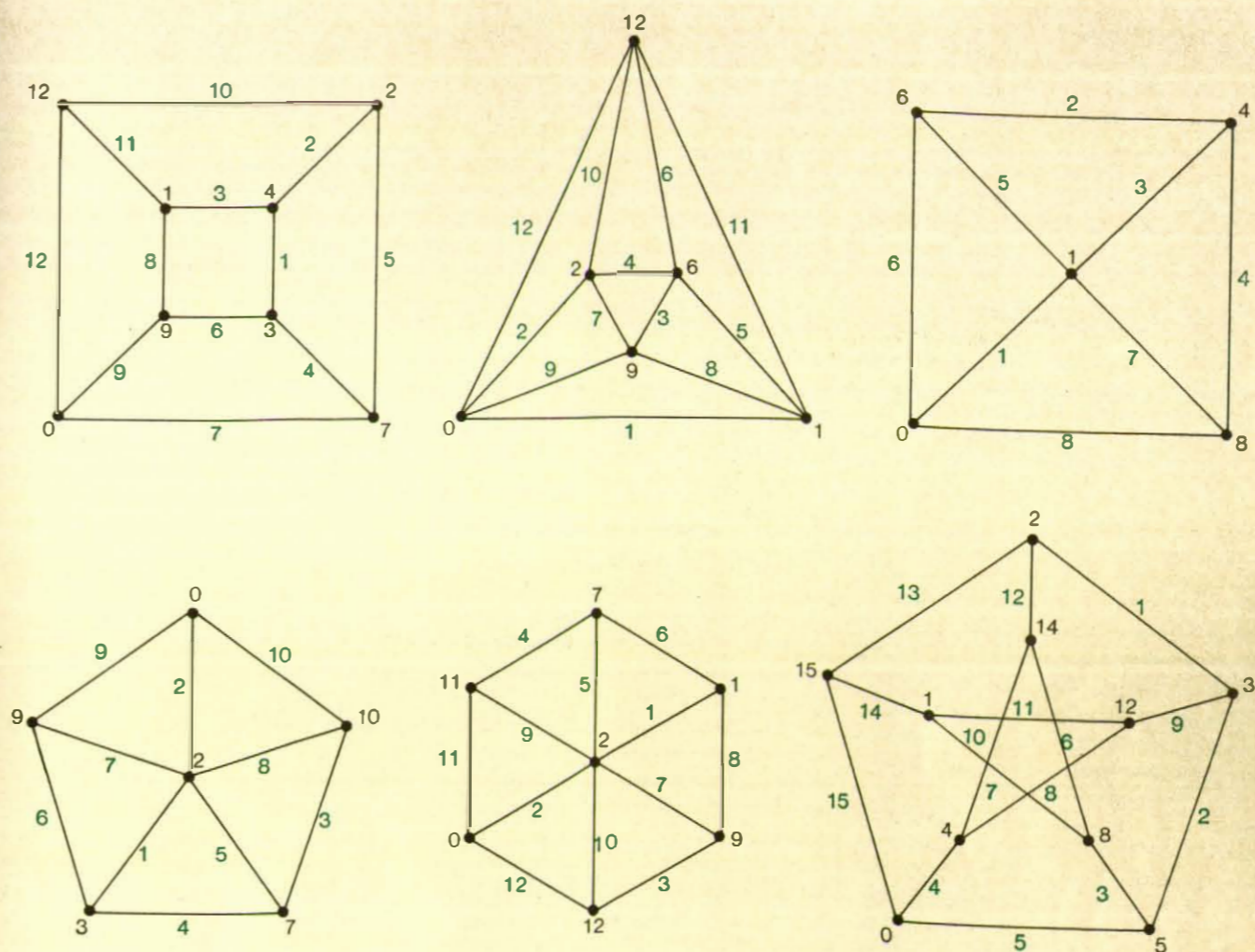
Trasferimento ortogonale delle pedine su una scacchiera 6 × 6.

Sei problemi riconducibili al gioco dell'alma.

re il problema sulla scacchiera normale di otto caselle di lato. Che almeno 12 mosse siano necessarie è evidente da una semplice analogia. Le sei pedine della colonna 1 e della colonna 3 saltando possono arrivare solo alla colonna 7; tre di esse devono perciò fare almeno una mossa di un passo. Ho cercato invano per settimane di trovare una soluzione in 12 mosse finché Donald E. Knuth, un matematico della Stanford University, mi venne in soccorso riuscendo a dimostrarne l'impossibilità. La dimostrazione è troppo complicata per riportarla qui; dirò solo che si fonda sulla necessità per una delle pedine dei quattro angoli originari di passare all'altro colore, sul fatto che il « rovescio » di una soluzione è una soluzione diversa e altre considerazioni. I lettori possono dilettarsi a trovare soluzioni che comportino un numero minimo di mosse per i sei problemi di trasferimento. Nessuna di queste soluzioni sarà data il mese prossimo. Ma se qualcuno riuscirà ad abbassare il numero minimo di mosse noto per qualcuno dei tre problemi di trasferimento in diagonale, oppure riuscirà a dimostrare che i numeri dati sono veramente i numeri minimi possibili, ne pubblicherò i risultati in uno dei prossimi numeri.

La seconda famiglia di rompicapo suggerita dal problema di Matsuda trae origine dall'eliminazione di tutte le pedine saltate. Obiettivo del gioco è eliminare tutte le pedine tranne una, la quale dovrà raggiungere una casella specificamente designata, il tutto in un numero minimo di mosse secondo le regole dell'alma. Questi problemi sono simili a quelli del classico gioco solitario del piolo, se si prescinde dal fatto che la maggiore libertà di movimento consente soluzioni molto piú brevi e che risulta di solito molto piú difficile dimostrare che le soluzioni trovate rappresentino il minimo numero di mosse.

Consideriamo, per esempio, il rompicapo su una scacchiera di cinque caselle per lato, pubblicato per la prima volta nel 1908 da Sam Lloyd (*si veda « a » nella figura in questa pagina*). Egli assegnò a ogni pedina il nome di un possibile candidato per le elezioni presidenziali di quell'anno. L'idea era di eliminare otto uomini, lasciando il favorito nella casella centrale. Le mosse lecite erano quelle del gioco dell'alma, con l'unica eccezione che non erano consentite serie di salti ma solo salti singoli. Otto salti sono evidentemente il numero minimo e per ogni pedina esistono molte soluzioni del genere. Henry Ernest Dudeney, in *Amusements in Mathematics*, perfezionò il rompicapo non consentendo passi, considerando una serie di salti connessi una sola mossa e consentendo a ciascuna pedina di finire al centro. Egli diede una soluzione in quattro mosse che è sicuramente minima, benché io non ne conosca alcuna dimostrazione. La pedina 5 salta la 8, la 9, la 3 e la 1; la pedina 7 salta la 4; la pedina 6 salta la 2 e la 7; e infine la pedina 5 torna alla sua casella di partenza saltando la 6.

Combiniamo ora le regole dei due enigmisti rivali consentendo passi e salti, come nell'alma, e considerando una serie di salti una sola mossa. Ogni pedina saltata viene ovviamente eliminata. Il lettore è in grado di trovare una tra le molte soluzioni in tre mosse che lasciano l'ultima pedina nella casella centrale? La soluzione che sarà data nel prossimo numero è una soluzione

La soluzione al problema della numerazione graziosa.

elegante che comincia con due passi e termina con una serie di otto salti.

Problemi simili sono illustrati in *b*, *c*, *d*, *e* e *f*. Di tutti sarà data la soluzione il mese prossimo: *b* dev'essere risolto in tre mosse e la pedina superstite dovrà trovarsi nella casella centrale; *c* in tre mosse, e l'ultima pedina dovrà trovarsi nella casella occupata dalla pedina 1; *d* in un numero minimo di mosse, e l'ultima pedina dovrà trovarsi nella casella centrale del triangolo; *e* in tre mosse, con la pedina superstite al centro. Il problema *f*, il piú difficile di tutti, richiede tre mosse, con la pedina finale in una delle quattro caselle centrali.

Pubblicherò le migliori soluzioni trovate dai lettori per questi sei problemi o per qualsiasi altro problema del tipo alma che presenti un interesse particolare. Il campo è cosí inesplorato che sfida a trovare e a risolvere nuovi problemi e a indagare la possibilità di dimostrare con argomentazioni semplici che la soluzione trovata è effettivamente quella che comporta il minimo numero di mosse possibile. Non ho, per esempio, la minima idea di quale sia il numero minimo di mosse richiesto, su una scacchiera con sette caselle di lato e con 25 pedine disposte in un quadrato al centro, per lasciare l'ultima pedina sulla casella centrale. Ho evitato di indagare questo problema per timore di non riuscire a far altro.

Le soluzioni al problema di numerare « graziosamente » i sei grafi proposti il mese scorso sono illustrati nella figura qui sopra. Nessuna delle numerazioni è unica. Se qualche lettore riuscirà a numerare graziosamente gli scheletri del dodecaedro e dell'icosaedro, ne pubblicherò la soluzione.

Nel numero del febbraio scorso ho commesso un errore riferendo che John L. Selfridge ha una dimostrazione sulla sua asserzione a proposito della regola della forchetta nel « go muku ». L'asserzione, egli mi ha successivamente informato, rimane una congettura. Egli riferisce nondimeno che una soluzione è stata trovata per il suo gioco del filetto su una scacchiera « quattro per infinito ». Questo gioco si pratica su una striscia che ha un'altezza di quattro caselle e una larghezza infinita; vince chi riesce a far arrivare per primo quattro delle sue pedine in una riga ortogonale o diagonale. Carlyle Lustenberger, nella sua tesi di laurea in scienza dei calcolatori all'Università della Pennsylvania, ha sviluppato un programma per calcolatori con una strategia vincente per il primo giocatore con una scacchiera 4 × 30. Il limite reale piú basso per la larghezza è inferiore di poche caselle, ma non ho ancora avuto particolari.

La scacchiera 3 × infinito consente una vittoria banale per il primo giocatore alla sua terza mossa; di fatto, la stessa vittoria può essere conseguita aggiungendo una sola casella al lato o una casella d'angolo al campo tradizionale del gioco del filetto su una scacchiera 3 × infinito. Il problema della scacchiera 5 × infinito non ha ancora trovato una soluzione. Se si potesse trovare un modo di vincere, su questa scacchiera, per il primo giocatore, la soluzione potrebbe essere naturalmente utilizzata anche per il go muku, quando viene giocato su un quadrato grande a piacere, senza regole restrittive.